

STIC-Biotech/ChemLib

180 507

From: Li, Bao-Qun
Sent: Tuesday, August 05, 2003 11:47 AM
To: STIC-Biotech/ChemLib

RECEIVED

AUG -5 2003

Please do the oligmer sequence search for SEQ ID NO: 1, 2, 5 and 6 for at least 80 % homology. The Application NO. is 10,074,620. Thank you.
CM1, 8E12.
AU 1648.

(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:44:31 ; Search time 32.25 Seconds

(without alignments)
273.726 Million cell updates/sec

Title: US-10-074-620-5

Sequence: 1 agsgatgcctgcacacaaga 20

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0 1139956

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 120 summaries

Database :

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCFUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	90.0	20	US-09-311-260-83	Sequence 83, Appl
2	16	80.0	2084	US-08-720-899-3	Sequence 3, Appl
3	16	80.0	2084	US-08-459-610-3	Sequence 3, Appl
4	16	80.0	2084	US-08-343-804-3	Sequence 3, Appl
5	16	80.0	2084	US-08-687-399-3	Sequence 3, Appl
6	16	80.0	2084	US-08-600-908A-3	Sequence 3, Appl
7	16	80.0	2084	US-08-683-838A-3	Sequence 3, Appl
8	16	80.0	2084	US-09-182-859-3	Sequence 3, Appl
9	16	80.0	2084	US-09-170-670-13	Sequence 13, Appl
10	16	80.0	2084	US-09-193-068-31	Sequence 31, Appl
11	16	80.0	2084	US-09-183-413-58	Sequence 58, Appl
12	16	80.0	2084	US-09-290-734-13	Sequence 13, Appl
13	16	80.0	2084	US-09-672-459-3	Sequence 3, Appl
14	16	80.0	2084	US-09-636-282A-3	Sequence 3, Appl
15	16	80.0	2084	US-09-545-586-13	Sequence 13, Appl
16	16	80.0	2084	US-09-264-097-3	Sequence 3, Appl
17	16	80.0	2604	US-09-537-168-5	Sequence 5, Appl
18	15	75.0	4403765	US-09-103-840A-2	Sequence 2, Appl
19	15	75.0	4411529	US-09-103-840A-1	Sequence 1, Appl
20	14	70.0	817	US-09-247-155-142	Sequence 142, Appl
21	13	65.0	20	US-07-977-289A-84	Sequence 84, Appl
22	13	65.0	20	US-08-236-426B-84	Sequence 84, Appl
23	13	65.0	922	US-09-609-816-2	Sequence 2, Appl
24	13	65.0	942	US-09-609-816-1	Sequence 1, Appl
25	13	65.0	1000	US-09-018-584A-41	Sequence 41, Appl
26	13	65.0	1004	US-09-609-816-3	Sequence 3, Appl
27	13	65.0	2045	US-09-152-060-22	Sequence 22, Appl

101 12 60.0 926 3 US-09-344-899-1 Sequence 1, Appli
c 102 12 60.0 944 3 US-09-070-526-1 Sequence 1, Appli
103 12 60.0 948 4 US-09-252-991A-10132 Sequence 10132, A
c 104 12 60.0 994 3 US-09-008-271A-19 Sequence 19, Appl
c 105 12 60.0 1047 4 US-09-252-991A-13653 Sequence 1653, A
c 106 12 60.0 1049 4 US-09-386-642-9 Sequence 9, Appli
107 12 60.0 1096 4 US-09-716-161A-3 Sequence 3, Appli
108 12 60.0 1101 3 US-08-746-883-3 Sequence 3, Appli
c 109 12 60.0 1112 4 US-09-716-161A-10 Sequence 10, Appl
c 110 12 60.0 1163 4 US-09-620-312D-250 Sequence 5, Appl
c 111 12 60.0 1212 2 US-08-449-933-5 Sequence 250, App
c 112 12 60.0 1212 3 US-07-966-049A-5 Sequence 5, Appli
c 113 12 60.0 1221 3 US-08-965-600-2 Sequence 2, Appli
114 12 60.0 1221 4 US-09-489-506-2 Sequence 2, Appli
115 12 60.0 1245 4 US-09-996-243-236 Sequence 236, App
c 116 12 60.0 1245 4 US-09-252-991A-13463 Sequence 13463, A
c 117 12 60.0 1317 4 US-09-252-991A-9768 Sequence 9768, App
118 12 60.0 1326 4 US-09-620-312D-867 Sequence 867, Appl
119 12 60.0 1386 1 US-08-672-571A-4 Sequence 4, Appli
c 120 12 60.0 1430 2 US-08-549-004A-15 Sequence 15, Appli

ALIGNMENTS

RESULT 1
US-09-311-260-83
Sequence 83, Application US/09311260
Patent No. 6234555
GENERAL INFORMATION:
APPLICANT: Leushner, James
APPLICANT: Hul, May
APPLICANT: Lacroix, Jean-Michel
TITLE OF INVENTION: METHOD, COMPOSITIONS AND KIT FOR DETECTION OF
TITLE OF INVENTION: MICROORGANISMS AND BI-DIRECTIONAL SEQUENCING OF NUCLEIC ACID
TITLE OF INVENTION: POLYMERS
NUMBER OF SEQUENCES: 189
CORRESPONDENCE ADDRESS:
ADDRESSEE: Opedah, & Larson LLP
STREET: P.O. Box 5270
CITY: Frisco
STATE: CO
COUNTRY: US
ZIP: 80443-5270
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS DOS
SOFTWARE: Word Perfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/311,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Larson, Marina T.
REGISTRATION NUMBER: 32,038
REFERENCE/DOCKET NUMBER: VGEN-P-058-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (970) 668-2050
TELEFAX: (970) 668-2082
TELEX:
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 20
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
HYPOTHETICAL: no

ANTI-SENSE: yes
FRAGMENT TYPE: internal
US-09-311-260-83

Query Match 90.0% Score 18; DB 3; Length 20;
Best Local Similarity 100.0% Fred. No. 0.046;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAGA 20
|||||
Db 3 GGATGCTGGACACAGA 20

RESULT 2
US-08-720-899-3/c
Sequence 3, Application US/08720899
Patent No. 5753460
GENERAL INFORMATION:
APPLICANT: Bisgaard-Frantzen, Henrik
APPLICANT: Borchert, Torben Vedel
APPLICANT: Svendsen, Allan
APPLICANT: Thellersen, Mariatne
APPLICANT: Van der Zee, Pia
TITLE OF INVENTION: AMYLASE VARIANTS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5753460c No. 5753460dix of No. 5753460th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/720,899
FILING DATE: 10-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/343,804
FILING DATE: 22-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Lowmey Dr., Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4054-214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-667-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2084 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 250..1794
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 250..342
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 343..1791
US-08-720-899-3
Query Match 80.0% Score 16; DB 1; Length 2084;
Best Local Similarity 100.0% Fred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGGACACAGA 18

Db 1292 GGATGCTGGACACAA 1277

RESULT 3
US-08-459-610-3/C

Sequence 3, Application US/08459610

Patent No. 5801043

GENERAL INFORMATION:

APPLICANT: Bisgaard-Frantzen, Henrik

APPLICANT: Borchert, Torben Vedel

APPLICANT: Svendsen, Allan

APPLICANT: Thellersen, Marianne

APPLICANT: Van der Zee, Pia

TITLE OF INVENTION: AMYLASE VARIANTS

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESSES:

ADDRESSEE: No. 58010430 No. 5801043disk of No. 5801043th America, Inc.

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/459,610

FILING DATE: 02-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/343,804

FILING DATE: 22-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Lowney Dr., Karen A.

REGISTRATION NUMBER: 31,274

REFERENCE/DOCKET NUMBER: 4054,214-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2084 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 250..1794

FEATURE:

NAME/KEY: sig_peptide

LOCATION: 250..342

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 343..1791

US-08-459-610-3

Query Match

Best Local Similarity 80.0%; Score 16; DB 1; Length 2084;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1292 GGATGCTGGACACAA 1277

QY 3 GGATGCTGGACACAA 18

Db 1292 GGATGCTGGACACAA 1277

RESULT 4

US-08-343-804-3/C

Sequence 3, Application US/08343804

Patent No. 5830837

GENERAL INFORMATION:

APPLICANT: Bisgaard-Frantzen, Henrik

APPLICANT: Borchert, Torben Vedel

APPLICANT: Svendsen, Allan

APPLICANT: Thellersen, Marianne

APPLICANT: Van der Zee, Pia

TITLE OF INVENTION: AMYLASE VARIANTS

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESSES:

ADDRESSEE: No. 58308370 No. 5830837disk of No. 5830837th America, Inc.

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/343,804

FILING DATE: 22-NOV-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lowney Dr., Karen A.

REGISTRATION NUMBER: 31,274

REFERENCE/DOCKET NUMBER: 4054,214-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123

TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2084 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 250..1794

FEATURE:

NAME/KEY: sig_peptide

LOCATION: 250..342

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 343..1791

US-08-343-804-3

Query Match

Best Local Similarity 80.0%; Score 16; DB 2; Length 2084;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1292 GGATGCTGGACACAA 1277

QY 3 GGATGCTGGACACAA 18

Db 1292 GGATGCTGGACACAA 1277

RESULT 5

US-08-687-399-3/C

Sequence 3, Application US/08687399

Patent No. 5928381

GENERAL INFORMATION:

APPLICANT: Toft, Annette H.

APPLICANT: Marcher, Dorte

APPLICANT: Pedersen, Hanne H.

APPLICANT: Nilsen, Thomas E.

TITLE OF INVENTION: A Combined Desizing and Bleaching

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: No. 59283810 No. 5928381disk of No. 5928381th America, Inc.

STREET: 405 Lexington Avenue, 64th Floor

CITY: New York

STATE: New York

COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin-Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,399
FILING DATE:
CLASSIFICATION: 008
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4127,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2084 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 250..342
FEATURE:
NAME/KEY: CDS
LOCATION: 250..1791
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 343..1791
US-08-687-399-3

Query Match 80.0%; Score 16; DB 2; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 6
US-08-600-908A-3/c
Sequence 3, Application US/08600908A
Patent No. 5989169
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Borchert, Torden Vedel
TITLE OF INVENTION: '-Amylase Mutants
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59891690 No. 5989169disk of No. 5989169th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/600,908A
FILING DATE: 13-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza

REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4394,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2084 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 250..1791
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 250..342
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 343..1791
US-08-600-908A-3

Query Match 80.0%; Score 16; DB 2; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 7
US-08-683-838A-3/c
Sequence 3, Application US/08683838A
Patent No. 6022724
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Bisg rd-Frantzen, Henrik
APPLICANT: Borchert, Torden Vedel
TITLE OF INVENTION: '-Amylase Mutants
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 60227240 No. 6022724disk of No. 6022724th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/683,838A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/600,908
FILING DATE: 13-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4394,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-9655
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2084 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 250..1791
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 250..342
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 343..1791
US-06-683-838A-3

Query Match 80.0%; Score 16; DB 3; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
DB 1292 GGATGCTGGACACAA 1277

RESULT 8
US-09-182-859-3/C
Sequence 3, Application US/09182859
Patent No. 6143708
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Borchert, Torben
APPLICANT: Bisgaard-Frantzen, Henrik
TITLE OF INVENTION: Alpha-Amylase Mutants
FILE REFERENCE: 4796.204-US
CURRENT APPLICATION NUMBER: US/09/182,859
EARLIER FILING DATE: 1998-10-29/96
EARLIER APPLICATION NUMBER: 0515/96
EARLIER FILING DATE: 1996-04-30
EARLIER APPLICATION NUMBER: 0712/96
EARLIER FILING DATE: 1996-06-28
EARLIER APPLICATION NUMBER: 0775/96
EARLIER FILING DATE: 1996-07-11
EARLIER APPLICATION NUMBER: 1263/96
EARLIER FILING DATE: 1996-11-08
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 2084
TYPE: DNA
ORGANISM: Bacillus amyloliquefaciens
US-09-182-859-3

Query Match 80.0%; Score 16; DB 3; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
DB 1292 GGATGCTGGACACAA 1277

RESULT 9
US-09-170-670-13/C
Sequence 13, Application US/09170670
Patent No. 6187576
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Borchert, Torben
APPLICANT: Bisgaard-Frantzen, Henrik
TITLE OF INVENTION: Alpha-Amylase Mutants
FILE REFERENCE: 5276.200-US
CURRENT APPLICATION NUMBER: US/09/170,670
EARLIER FILING DATE: 1998-10-13
EARLIER APPLICATION NUMBER: 1172/97
EARLIER FILING DATE: 1997-10-13

EARLIER APPLICATION NUMBER: 60/063,306
EARLIER FILING DATE: 1997-10-28
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 13
LENGTH: 2084
TYPE: DNA
ORGANISM: B. amyloliquefaciens
US-09-170-670-13

Query Match 80.0%; Score 16; DB 3; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
DB 1292 GGATGCTGGACACAA 1277

RESULT 10
US-09-193-068-31/C
Sequence 31, Application US/09193068
Patent No. 6197565
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Kjullif, S ren
APPLICANT: Bisgaard-Frantzen, Henrik
APPLICANT: Andersen, Carsten
TITLE OF INVENTION: -Amylase Variants
FILE REFERENCE: 5709.000-US
CURRENT APPLICATION NUMBER: US/09/193,068
EARLIER FILING DATE: 1998-11-16
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 31
LENGTH: 2084
TYPE: DNA
ORGANISM: B. amyloliquefaciens
US-09-193-068-31

Query Match 80.0%; Score 16; DB 3; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
DB 1292 GGATGCTGGACACAA 1277

RESULT 11
US-09-183-412-58/C
Sequence 58, Application US/09183412
Patent No. 6204232
GENERAL INFORMATION:
APPLICANT: Borchert, Torben V.
APPLICANT: Svendsen, Allan
APPLICANT: Andersen, Carsten
APPLICANT: Nielsen, Bjarne
APPLICANT: Nissen, Torben L.
APPLICANT: Kjaerulff, Soren
TITLE OF INVENTION: Alpha-Amylase Mutants
FILE REFERENCE: 5368.200-US
CURRENT APPLICATION NUMBER: US/09/183,412
EARLIER FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: 60/064,662
EARLIER FILING DATE: 1997-11-06
EARLIER APPLICATION NUMBER: 60/093,234
EARLIER FILING DATE: 1998-07-17
EARLIER APPLICATION NUMBER: 1240/97
EARLIER FILING DATE: 1997-10-30
EARLIER APPLICATION NUMBER: PA 1998 00936
EARLIER FILING DATE: 1998-07-14
NUMBER OF SEQ ID NOS: 58

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 58
LENGTH: 2084
TYPE: DNA
ORGANISM: Bacillus amyloliquefaciens
FEATURE:
NAME/KEY: CDS
LOCATION: (343)...(1794)
US-09-183-412-58

Query Match 80.0%; Score 16; DB 3; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTTGACACAA 18
|||||
DB 1292 GGATGCTTGACACAA 1277

RESULT 12
US-09-290-734-13/c
Sequence 13, Application US/09290734
Patent No. 6361989
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Borchert, Torben Vedel
APPLICANT: Bisgard-Frantzen Henrik
APPLICANT: Outtrup, Helge
APPLICANT: Nielsen, Bjarne Ronfeldt
APPLICANT: Nielsen, Vibeke Skovgaard
APPLICANT: Hoeck, Lisbeth Hedegaard
TITLE OF INVENTION: No. 6361989-1 -Amylase And -Amylase Mutants
FILE REFERENCE: 5276.400-US
CURRENT APPLICATION NUMBER: US/09/290.734
CURRENT FILING DATE: 1999-04-13
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 13
LENGTH: 2084
TYPE: DNA
ORGANISM: B. amyloliquefaciens
US-09-290-734-13

Query Match 80.0%; Score 16; DB 4; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTTGACACAA 18
|||||
DB 1292 GGATGCTTGACACAA 1277

RESULT 13
US-09-672-459-3/c
Sequence 3, Application US/09672459
Patent No. 6436888
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Borchert, Torben
APPLICANT: Bisgard-Frantzen, Henrik
TITLE OF INVENTION: Alpha-Amylase Mutants
FILE REFERENCE: 4796.204-US
CURRENT APPLICATION NUMBER: US/09/672.459
CURRENT FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: 09/182.859
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: 0515/96
PRIOR FILING DATE: 1996-04-30
PRIOR APPLICATION NUMBER: 0712/96
PRIOR FILING DATE: 1996-06-28
PRIOR APPLICATION NUMBER: 0775/96
PRIOR FILING DATE: 1996-07-11
PRIOR APPLICATION NUMBER: 1263/96

PRIOR FILING DATE: 1996-11-08
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 2084
TYPE: DNA
ORGANISM: Bacillus amyloliquefaciens
US-09-672-459-3

Query Match 80.0%; Score 16; DB 4; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTTGACACAA 18
|||||
DB 1292 GGATGCTTGACACAA 1277

RESULT 14
US-09-636-252A-3/c
Sequence 3, Application US/09636252A
Patent No. 6440716
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Bisgard-Frantzen, Henrik
APPLICANT: Borchert, Torben Vedel
APPLICANT: Outtrup, Helge
TITLE OF INVENTION: Alpha-Amylase Mutants
FILE REFERENCE: 0776/1P216-US2
CURRENT APPLICATION NUMBER: US/09/636.252A
CURRENT FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: 08/683.838
PRIOR FILING DATE: 1996-07-18
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 2084
TYPE: DNA
ORGANISM: B. amyloliquefaciens
FEATURE:
NAME/KEY: CDS
LOCATION: (250)...(1791)
NAME/KEY: mat_peptide
LOCATION: (343)...(1791)
NAME/KEY: sig_peptide
LOCATION: (250)...(342)
US-09-636-252A-3

Query Match 80.0%; Score 16; DB 4; Length 2084;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTTGACACAA 18
|||||
DB 1292 GGATGCTTGACACAA 1277

RESULT 15
US-09-545-586-13/c
Sequence 13, Application US/09545586
Patent No. 6528298
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
APPLICANT: Borchert, Torben Vedel
APPLICANT: Bisgard-Frantzen Henrik
APPLICANT: Outtrup, Helge
APPLICANT: Nielsen, Bjarne Ronfeldt
APPLICANT: Nielsen, Vibeke Skovgaard
APPLICANT: Hoeck, Lisbeth Hedegaard
TITLE OF INVENTION: No. 6528298-1 -Amylase And -Amylase Mutants
FILE REFERENCE: 5276.400-US
CURRENT APPLICATION NUMBER: US/09/545.586
CURRENT FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: US/09/290.734

PRIOR FILING DATE: 1999-04-13
 NUMBER OF SEQ ID NOS: 35
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 13
 LENGTH: 2084
 TYPE: DNA
 ORGANISM: B. amyloliquefaciens
 US-09-545-586-13

Query Match
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
 DB 1292 GGATGCTGGACACAA 1277

RESULT 16
 US-09-264-097-3/C
 Sequence 3, Application US/09264097
 Patent No. 6287826
 GENERAL INFORMATION:
 APPLICANT: No. 6287826man, Barrie Edmund
 APPLICANT: Hendriksen, Hanne Vang
 TITLE OF INVENTION: Enzymatic Preparation of Glucose Syrup
 FILE REFERENCE: 5278-200-US
 CURRENT APPLICATION NUMBER: US/09/264,097
 EARLIER FILING DATE: 1999-03-08
 EARLIER APPLICATION NUMBER: PA 0321/98
 EARLIER FILING DATE: 1998-03-09
 EARLIER APPLICATION NUMBER: 60/079,209
 NUMBER OF SEQ ID NOS: 8
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 3
 LENGTH: 2604
 TYPE: DNA
 ORGANISM: Bacillus amyloliquefaciens
 US-09-264-097-3

Query Match
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
 DB 1812 GGATGCTGGACACAA 1797

RESULT 17
 US-09-537-168-5/C
 Sequence 5, Application US/09537168
 Patent No. 6410295
 GENERAL INFORMATION:
 APPLICANT: Andersen, Carsten
 APPLICANT: Jorgensen, Christel T.
 APPLICANT: Bisgaard-Frantzen, Henrik
 APPLICANT: Svendsen, Allan
 APPLICANT: Kjaerulff, Soren
 TITLE OF INVENTION: Alpha-Amylase Variants
 FILE REFERENCE: 5886-200-US
 CURRENT APPLICATION NUMBER: US/09/537,168
 EARLIER FILING DATE: 2000-03-29
 EARLIER APPLICATION NUMBER: PA 1999 00437
 EARLIER FILING DATE: 1999-03-30
 EARLIER APPLICATION NUMBER: 60/127,427
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 5
 LENGTH: 2604

TYPE: DNA
 ORGANISM: Bacillus amyloliquefaciens
 US-09-537-168-5

Query Match
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
 DB 1812 GGATGCTGGACACAA 1797

RESULT 18
 US-09-103-840A-2/C
 Sequence 2, Application US/09103840A
 Patent No. 6294328
 GENERAL INFORMATION:
 APPLICANT: FLEISCHMAN, Robert D.
 APPLICANT: WHITE, Owen R.
 APPLICANT: FRASER, Claire M.
 APPLICANT: VENTER, John C.
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 FILE REFERENCE: 24366-20007.00
 CURRENT APPLICATION NUMBER: US/09/103,840A
 EARLIER FILING DATE: 1998-06-24
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2
 LENGTH: 4403765
 TYPE: DNA
 ORGANISM: Mycobacterium tuberculosis
 FEATURE:
 OTHER INFORMATION: CDC 1551
 OTHER INFORMATION: "n" bases at various positions throughout the sequence
 OTHER INFORMATION: represent a, c, g or g
 US-09-103-840A-2

Query Match
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGGACAA 15
 DB 2965084 AGGATGCTGGACAA 2965070

RESULT 19
 US-09-103-840A-1/C
 Sequence 1, Application US/09103840A
 Patent No. 6294328
 GENERAL INFORMATION:
 APPLICANT: FLEISCHMAN, Robert D.
 APPLICANT: WHITE, Owen R.
 APPLICANT: FRASER, Claire M.
 APPLICANT: VENTER, John C.
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 FILE REFERENCE: 24366-20007.00
 CURRENT APPLICATION NUMBER: US/09/103,840A
 EARLIER FILING DATE: 1998-06-24
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 1
 LENGTH: 4411529
 TYPE: DNA
 ORGANISM: Mycobacterium tuberculosis
 OTHER INFORMATION: H37Rv
 US-09-103-840A-1

Query Match
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGGACA 15
DB 2968978 AGGATGCTGGACA 2968964

RESULT 20
US-09-247-155-142/c

Sequence 142, Application US/09247155A

Patent No. 6312822

GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, Jean-Baptiste

APPLICANT: Bouquelere, Lydie

TITLE OF INVENTION: Complementary DNAs

FILE REFERENCE: GENSET.021A

CURRENT APPLICATION NUMBER: US/09/247,155A

EARLIER FILING DATE: 1998-02-09

EARLIER FILING DATE: 1998-02-09

EARLIER FILING DATE: 1998-04-13

EARLIER FILING DATE: 1998-06-11

EARLIER FILING DATE: 1998-08-10

EARLIER FILING DATE: 1998-10-04

NUMBER OF SEQ ID NOS: 182

SOFTWARE: Patent.Pm

SEQ ID NO 142

LENGTH: 817

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: CDS

LOCATION: 28..804

FEATURE:
NAME/KEY: sig_peptide

LOCATION: 28..96

OTHER INFORMATION: Von Heijne matrix

OTHER INFORMATION: score 10

OTHER INFORMATION: seq FULGILLSPPAGA/DV

FEATURE:
NAME/KEY: polyA_site

LOCATION: 806..817

US-09-247-155-142

Query Match

Best Local Similarity 70.0%; Score 14; DB 4; Length 817;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ATGCTGGACACAA 18
DB 333 ATGCTGGACACAA 320

RESULT 21
US-07-977-284A-84

Sequence 84, Application US/07977284A

Patent No. 5558988

GENERAL INFORMATION:
APPLICANT: Prockop, Darwin J.

APPLICANT: Ala-Kokko, Leena

APPLICANT: Williams, Charlene J.

APPLICANT: Rivtaniemi, Pertti

APPLICANT: Baldwin, Clinton

APPLICANT: Hopkinson, Ian

APPLICANT: Ahmad, Nilofar Nina

TITLE OF INVENTION: METHODS OF DETECTING A GENETIC

TITLE OF INVENTION: PREDISPOSITION FOR OSTEOARTHRITIS

NUMBER OF SEQUENCES: 261

CORRESPONDENCE ADDRESS:
ADDRESS: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5558988r1s

STREET: One Liberty Place, 46th floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/977,284A

FILING DATE: 13-NOV-1992

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Deluca, Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-0697

TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 84:

SEQUENCE CHARACTERISTICS:

LENGTH: 20

TYPE: NUCLEIC ACID

STRANDEDNESS: SINGLE

TOPOLOGY: LINEAR

ANTI-SENSE: YES

US-07-977-284A-84

Query Match

Best Local Similarity 65.0%; Score 13; DB 1; Length 20;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GATGCTGGACAC 16
DB 5 GATGCTGGACAC 17

RESULT 22
US-08-256-426B-84

Sequence 84, Application US/08256426B

Patent No. 5948611

GENERAL INFORMATION:
APPLICANT: Prockop, Darwin J.

APPLICANT: Ala-Kokko, Leena

APPLICANT: Williams, Charlene J.

APPLICANT: Rivtaniemi, Pertti

APPLICANT: Baldwin, Clinton

APPLICANT: Hopkinson, Ian

APPLICANT: Ahmad, Nilofar Nina

TITLE OF INVENTION: Methods of Detecting A Genetic

NUMBER OF SEQUENCES: 293

CORRESPONDENCE ADDRESS:
ADDRESS: Woodcock, Washburn Kurtz Mackiewicz & No. 5948611r1s

STREET: One Liberty Place - 46th floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.5 INCH

COMPUTER: IBM compatible

OPERATING SYSTEM: Windows 3.1

SOFTWARE: WORDPERFECT 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/256,426B

FILING DATE: 03-FEB-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/10964
FILING DATE: 12-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,284
FILING DATE: 13-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: MARK DELUCA
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1082
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 20
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
ANTI-SENSE: YES
US-08-256-42GB-84

Query Match 65.0%; Score 13; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GATGCTGACAC² 16
DB 5 GATGCTGACAC 17

RESULT 23
US-09-609-816-2
Sequence 2, Application US/09609816
Patent No. 6436684
GENERAL INFORMATION:
APPLICANT: Woodage, Trevor
APPLICANT: Wei, Minh Hui
APPLICANT: Kodira, Chinappa
APPLICANT: Beasley, Ellen
APPLICANT: DiFrancesco, Valentina
TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
FILE REFERENCE: CL000669PCT
CURRENT APPLICATION NUMBER: US/09/609,816
PRIOR FILING DATE: 2000-07-03
PRIOR APPLICATION NUMBER: 60/192,408
PRIOR FILING DATE: 2000-03-27
PRIOR APPLICATION NUMBER: 60/212,725
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/609,816
PRIOR FILING DATE: 2000-07-03
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 922
TYPE: DNA
ORGANISM: HUMAN
US-09-609-816-2

Query Match 65.0%; Score 13; DB 4; Length 922;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGGA 13
DB 808 AGGATGCTGGA 820

RESULT 24
US-09-609-816-1
Sequence 1, Application US/09609816
Patent No. 6436684

GENERAL INFORMATION:
APPLICANT: Woodage, Trevor
APPLICANT: Wei, Minh Hui
APPLICANT: Kodira, Chinappa
APPLICANT: Beasley, Ellen
APPLICANT: DiFrancesco, Valentina
TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
FILE REFERENCE: CL000669PCT
CURRENT APPLICATION NUMBER: US/09/609,816
PRIOR FILING DATE: 2000-07-03
PRIOR APPLICATION NUMBER: 60/192,408
PRIOR FILING DATE: 2000-03-27
PRIOR APPLICATION NUMBER: 60/212,725
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/609,816
PRIOR FILING DATE: 2000-07-03
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 942
TYPE: DNA
ORGANISM: HUMAN
US-09-609-816-1

Query Match 65.0%; Score 13; DB 4; Length 942;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGGA 13
DB 808 AGGATGCTGGA 820

RESULT 25
US-09-018-584A-41/C
Sequence 41, Application US/09018584A
Patent No. 6238863
GENERAL INFORMATION:
APPLICANT: Schumm, James W.
APPLICANT: Bachet, Jeffrey W.
TITLE OF INVENTION: MATERIALS AND METHODS FOR
TITLE OF INVENTION: IDENTIFYING AND ANALYZING INTERMEDIATE TANDEM
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Promega Corporation
STREET: 2800 Woods Hollow Road
CITY: Madison
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53711-5399
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb
COMPUTER: IBM compatible PC
OPERATING SYSTEM: Windows 95
SOFTWARE: Word 97 (DOS text format)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,584A
FILING DATE: 04-Feb-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Grady J. Frenchick
REGISTRATION NUMBER: 29,018
REFERENCE/DOCKET NUMBER: 16026.9180
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 257-3501
TELEFAX: (608) 257-2275
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 1000 bp
TYPE: Nucleic Acid


```
STRANDEDNESS: Double
TOPOLOGY: Circular
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: no
IMMEDIATE SOURCE:
LIBRARY:
CLONE: S097
US-09-018-584A-41

Query Match
Best Local Similarity 100.0%; Score 13; DB 3; Length 1000;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTTGA 13
DB 445 AGGATGCTTGA 433

RESULT 26
US-09-609-816-3
Sequence 3, Application US/09609816
Patent No. 6436684
GENERAL INFORMATION:
APPLICANT: Woodage, Trevor
APPLICANT: Wei, Minh Hui
APPLICANT: Kodira, Chiappa
APPLICANT: Beasley, Ellen
APPLICANT: DiFrancesco, Valentina
TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
FILE REFERENCE: C1000669PCT
CURRENT APPLICATION NUMBER: US/09/609,816
PRIOR FILING DATE: 2000-07-03
PRIOR APPLICATION NUMBER: 60/192,408
PRIOR FILING DATE: 2000-03-27
PRIOR APPLICATION NUMBER: 60/212,725
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/609,816
PRIOR FILING DATE: 2000-07-03
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1004
TYPE: DNA
ORGANISM: HUMAN
US-09-609-816-3

Query Match
Best Local Similarity 100.0%; Score 13; DB 4; Length 1004;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTTGA 13
DB 847 AGGATGCTTGA 859

RESULT 27
US-09-152-060-22/C
Sequence 22, Application US/09152060
Patent No. 6448230
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 28 Human Secreted Proteins
FILE REFERENCE: P2003P1.US
CURRENT APPLICATION NUMBER: US/09/152,060
CURRENT FILING DATE: 1998-09-11
EARLIER APPLICATION NUMBER: PCT/US98/04858
EARLIER FILING DATE: 1998-03-12
EARLIER APPLICATION NUMBER: 60/040,762
EARLIER FILING DATE: 1997-03-14
EARLIER APPLICATION NUMBER: 60/040,710
EARLIER FILING DATE: 1997-03-14
```

```
EARLIER APPLICATION NUMBER: 60/050,934
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,100
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,357
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,189
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/057,765
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/068,368
EARLIER FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 118
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 22
LENGTH: 2045
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2040)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (2041)
OTHER INFORMATION: n equals a,t,g, or c
US-09-152-060-22

Query Match
Best Local Similarity 100.0%; Score 13; DB 4; Length 2045;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CCTGACACAGA 20
DB 1494 CCTGACACAGA 1482

RESULT 28
US-09-620-312D-621/C
Sequence 621, Application US/09620312D
Patent No. 6569652
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Dantui
APPLICANT: Wang, Zhimei
APPLICANT: John Tillinghast
APPLICANT: Dmanac, Redoyle T.
TITLE OF INVENTION: No. 656962e1 Nucleic Acids and
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_FL_genes Version 1.0
SEQ ID NO 621
```

LENGTH: 2315
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (257)..(937)
 US-09-620-312D-621

Query Match 65.0% Score 13; DB 4; Length 2315;
 Best Local Similarity 100.0% Pred. No. 40;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 8 CCTGACACACAGA 20
 DB 1588 CCTGACACACAGA 1576

RESULT 29

US-09-149-476-227
 Sequence 227, Application US/09149476
 Patent No. 6420526
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: 186 Human Secreted proteins
 FILE REFERENCE: P2002P1
 CURRENT APPLICATION NUMBER: US/09/149,476
 EARLIER FILING DATE: 1998-09-08
 EARLIER APPLICATION NUMBER: PCT/US98/04493
 EARLIER FILING DATE: 1998-03-06
 EARLIER APPLICATION NUMBER: 60/040,162
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,333
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/038,621
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,626
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,334
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,336
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/040,163
 EARLIER FILING DATE: 1997-03-07
 EARLIER APPLICATION NUMBER: 60/047,600
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,615
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,597
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,502
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,633
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,583
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,617
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,618
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,503
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,592
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,581
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,584
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,500
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,587
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,492
 EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047,598
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,613
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,582
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,596
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,612
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,632
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,601
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/043,580
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,566
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,314
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,569
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,311
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,671
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,674
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,669
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,312
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,313
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,672
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,315
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/048,974
 EARLIER FILING DATE: 1997-06-06
 EARLIER APPLICATION NUMBER: 60/056,886
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,877
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,889
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,893
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,630
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,878
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,662
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,872
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,882
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,637
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,903
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,888
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,879
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,880
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,894
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,911
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,636

EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,874
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,910
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,864
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,631
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,845
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,892
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/057,761
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/047,595
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,599
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,588
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,585
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,586
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,590
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,594
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,589
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/043,593
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/047,614
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/043,578
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/043,576
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/047,501
 EARLIER FILING DATE: 1997-05-23
 EARLIER APPLICATION NUMBER: 60/043,670
 EARLIER FILING DATE: 1997-04-11
 EARLIER APPLICATION NUMBER: 60/056,632
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,664
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,876
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,881
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,909
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,875
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,862
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,887
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/056,908
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/048,964
 EARLIER FILING DATE: 1997-06-06
 EARLIER APPLICATION NUMBER: 60/057,650
 EARLIER FILING DATE: 1997-09-05
 EARLIER APPLICATION NUMBER: 60/056,884
 EARLIER FILING DATE: 1997-08-22
 EARLIER APPLICATION NUMBER: 60/057,669
 EARLIER FILING DATE: 1997-09-05
 EARLIER APPLICATION NUMBER: 60/049,610
 EARLIER FILING DATE: 1997-06-13
 EARLIER APPLICATION NUMBER: 60/061,060
 EARLIER FILING DATE: 1997-10-02

Query Match 65.0%; Score 13; DB 4; Length 2517;
 Best Local Similarity 100.0%; Pred. No. 40;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGATGCTGGA 13
 DB 1391 AGGATGCTGGA 1403

RESULT 30
 US-09-149-476-61
 ; Sequence 61, Application US/09149476
 ; Patent No. 6420526
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: 166 Human Secreted proteins
 ; FILE REFERENCE: P2002P1
 ; CURRENT APPLICATION NUMBER: US/09/149,476
 ; CURRENT FILING DATE: 1998-09-08
 ; EARLIER APPLICATION NUMBER: PCT/US98/04493
 ; EARLIER FILING DATE: 1998-03-06
 ; EARLIER APPLICATION NUMBER: 60/040,162
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,333
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/038,621
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,626
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,334
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,336
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/040,163
 ; EARLIER FILING DATE: 1997-03-07
 ; EARLIER APPLICATION NUMBER: 60/047,600
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,615
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,597
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,502
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,633
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,583
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,617
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,618
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,503
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,592
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,581
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,584
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,500
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,587
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,492
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,598
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,613
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,582
 ; EARLIER FILING DATE: 1997-05-23
 ; EARLIER APPLICATION NUMBER: 60/047,596

Mon Aug 18 10:30:09 2003

us-10-074-620-5.011.rnt

Page 13

EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056,631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Query Match 65.0%; Score 13; DB 4; Length 2888;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 AGGATCCCTGCA 13
|||||

Db 1733 AGGATGCTGGA 1745

RESULT 31

US-09-110-517-3/C
Sequence 3, Application US/09110517A
Patent No. 6248520
GENERAL INFORMATION:
APPLICANT: Roeder, Robert G
APPLICANT: Pondell, Joseph D
APPLICANT: Yuan, Chao X
APPLICANT: Ito, Mitsuhito
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NUCLEAR HORMONE
FILE REFERENCE: 600-1-224
CURRENT APPLICATION NUMBER: US/09/110, 517A
CURRENT FILING DATE: 1998-07-06
NUMBER OF SEQ ID NOS: 51
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 2970
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2970)
US-09-110-517-3

Query Match 65.0%; Score 13; DB 3; Length 2970;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ATGCTGACACA 17

Db 289 ATGCTGACACA 277

RESULT 32

US-08-592-126-85/C
Sequence 85, Application US/08592126
Patent No. 5621091

GENERAL INFORMATION:
APPLICANT: Gregory Dolgancv
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
TITLE OF INVENTION: Polypeptides
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA

ZIP: 94306
COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,126
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:

LENGTH: 2995 base pairs
TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: A116con.seq
US-08-592-126-85

Query Match 65.0%; Score 13; DB 1; Length 2995;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGACACA 15

Db 1366 GGATGCTGACACA 1354

RESULT 33

US-09-168-595-85/C
Sequence 85, Application US/09168595
Patent No. 655866

GENERAL INFORMATION:
APPLICANT: Gregory Dolgancv
TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
TITLE OF INVENTION: Polypeptides
NUMBER OF SEQUENCES: 151
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA

ZIP: 94306
COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/168,595
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/592,126
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615

REFERENCE/DOCKET NUMBER: 4600-0111
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960

INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:

LENGTH: 2995 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: A116con.seq
US-09-168-595-85

Query Match 65.0%; Score 13; DB 4; Length 2995;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGACACA 15

Db 1366 GGATGCTGACACA 1354

```
RESULT 34
US-09-220-132-26/C
; Sequence 26, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 26
; LENGTH: 3468
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-26

Query Match      65.0%; Score 13; DB 4; Length 3468;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 ATGCGCTGACACA 17
Db      362 ATGCGCTGACACA 350

RESULT 35
US-09-019-201A-1/C
; Sequence 1, Application US/09019201A
; Patent No. 5968780
; GENERAL INFORMATION:
; APPLICANT: SORPET, DANIEL R.
; APPLICANT: LI, YI
; APPLICANT: DILLON, PATRICK J.
; TITLE OF INVENTION: DENDRITIC CELL-DERIVED GROWTH FACTOR
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,201A
; FILING DATE: HERewith
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: A. ANDERS BROOKES
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF346
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3621 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 134..1666
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 134..212
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 215..1666
US-09-019-201A-1

Query Match      65.0%; Score 13; DB 2; Length 3621;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 GGGATGCGCTGAC 14
Db      132 GGGATGCGCTGAC 120
```

Search completed: August 15, 2003, 11:00:34
Job time : 46.25 secs

Mon Aug 18 10:30:10 2003

us-10-074-620-6.011.rng

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 20:57:44 ; Search time 138.6 Seconds
(without alignments)
428.482 Million cell updates/sec

Title: US-10-074-620-6
Perfect score: 22
Sequence: 1 attgcaccaccagcagcacc 22

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 120 summaries

Database : N_Geneseq_19jun03.*
1: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT.*
21: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2001.DAT.*
22: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT.*
23: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2003.DAT.*
24: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2003.DAT.*
25: /SIDSI/gcgcdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	AAH03060	Microorganism dete
2	22	100.0	22	ABA000273	ESNA 2 antisense p
3	19	86.4	22	ABA50785	Human foetal liver
4	19	86.4	22	ABA68749	Human foetal liver
5	19	86.4	22	ABA5715	Probe #14181 for g
6	19	86.4	22	AAK17094	Human brain expres
7	19	86.4	22	AAK42877	Human bone marrow
8	19	86.4	22	AAI23642	Probe #13575 for g

9	19	86.4	22	AAI48952	Probe #17638 used
10	19	86.4	22	AAI09257	Probe #9248 used t
11	19	86.4	22	AB542507	Human liver single
12	19	86.4	22	AB516932	Human genome-deriv
13	19	86.4	22	ABN45663	Human breast cell
14	19	86.4	22	ABN45616	Human foetal liver
15	19	86.4	22	ABN45616	Probe #4282 for ge
16	19	86.4	22	AAK04355	Human brain expres
17	19	86.4	22	AAK29849	Human bone marrow
18	19	86.4	22	AAI14440	Probe #4373 for ge
19	19	86.4	22	AAI14440	Probe #4498 used t
20	19	86.4	22	AAI04263	Probe #4254 used t
21	19	86.4	22	AB529496	Human liver single
22	19	86.4	22	AB504410	Human genome-deriv
23	19	86.4	22	ABL16705	Drosophila melanog
24	19	86.4	22	AAK61084	P. putida KT2440-a
25	19	86.4	22	AAK61084	P. putida KT2440-a
26	19	86.4	22	ABL16704	Drosophila melanog
27	18	81.8	22	AAK75404	Codon-optimised HP
28	18	81.8	22	ABK14577	Human papilloma vi
29	18	81.8	22	ABL29445	Drosophila melanog
30	18	81.8	22	AAH73845	Permutin protein
31	18	81.8	22	AAQ72603	Plant insecticidal
32	18	81.8	22	AAH73850	Permutin protein
33	18	81.8	22	AAH73853	Permutin protein
34	18	81.8	22	AAK75383	Codon-optimised HP
35	18	81.8	22	ABL58977	S. pneumoniae type
36	18	81.8	22	ABX06947	HPV16-11 encoding
37	18	81.8	22	ABL58982	HPV16-11 encoding
38	18	81.8	22	ABL58982	Drosophila melanog
39	18	81.8	22	ABL28523	Drosophila melanog
40	18	81.8	22	AAK29434	Human MET11 relate
41	18	81.8	22	AAK29434	113855 cDNA clone.
42	18	81.8	22	ABL28522	Drosophila melanog
43	18	81.8	22	AAK29434	Streptococcus pneu
44	18	81.8	22	AAK29434	Streptococcus pneu
45	17	77.3	22	AAK29434	Streptococcus pneu
46	17	77.3	22	AAK29434	Streptococcus pneu
47	17	77.3	22	AAK29434	Streptococcus pneu
48	17	77.3	22	AAK29434	Streptococcus pneu
49	17	77.3	22	AAK29434	Streptococcus pneu
50	17	77.3	22	AAK29434	Streptococcus pneu
51	17	77.3	22	AAK29434	Streptococcus pneu
52	17	77.3	22	AAK29434	Streptococcus pneu
53	17	77.3	22	AAK29434	Streptococcus pneu
54	17	77.3	22	AAK29434	Streptococcus pneu
55	17	77.3	22	AAK29434	Streptococcus pneu
56	17	77.3	22	AAK29434	Streptococcus pneu
57	17	77.3	22	AAK29434	Streptococcus pneu
58	17	77.3	22	AAK29434	Streptococcus pneu
59	17	77.3	22	AAK29434	Streptococcus pneu
60	17	77.3	22	AAK29434	Streptococcus pneu
61	17	77.3	22	AAK29434	Streptococcus pneu
62	17	77.3	22	AAK29434	Streptococcus pneu
63	17	77.3	22	AAK29434	Streptococcus pneu
64	17	77.3	22	AAK29434	Streptococcus pneu
65	17	77.3	22	AAK29434	Streptococcus pneu
66	17	77.3	22	AAK29434	Streptococcus pneu
67	17	77.3	22	AAK29434	Streptococcus pneu
68	17	77.3	22	AAK29434	Streptococcus pneu
69	17	77.3	22	AAK29434	Streptococcus pneu
70	17	77.3	22	AAK29434	Streptococcus pneu
71	17	77.3	22	AAK29434	Streptococcus pneu
72	17	77.3	22	AAK29434	Streptococcus pneu
73	17	77.3	22	AAK29434	Streptococcus pneu
74	17	77.3	22	AAK29434	Streptococcus pneu
75	17	77.3	22	AAK29434	Streptococcus pneu
76	17	77.3	22	AAK29434	Streptococcus pneu
77	17	77.3	22	AAK29434	Streptococcus pneu
78	17	77.3	22	AAK29434	Streptococcus pneu
79	17	77.3	22	AAK29434	Streptococcus pneu
80	17	77.3	22	AAK29434	Streptococcus pneu
81	17	77.3	22	AAK29434	Streptococcus pneu

C	82	17	77.3	1461	24	ABK10752	DNA encoding human
C	83	17	77.3	1718	24	ABO99548	Human coding sequ
C	84	17	77.3	1733	22	AA803059	Human diagnostic a
C	85	17	77.3	1776	23	AA82210	DNA encoding novel
C	86	17	77.3	1845	23	AA889517	DNA encoding novel
C	87	17	77.3	1879	23	AB105443	Drosophila melanog
C	88	17	77.3	2199	23	AB117267	Drosophila melanog
C	89	17	77.3	2219	23	AA139627	Human secreted pro
C	90	17	77.3	2829	23	AA590299	DNA encoding novel
C	91	17	77.3	2994	23	AB110625	Drosophila melanog
C	92	17	77.3	3028	23	AA922734	Chrysosporium Xyl
C	93	17	77.3	3028	24	AA772046	Chrysosporium Xyl
C	94	17	77.3	3561	24	AA597178	Human metalloprote
C	95	17	77.3	3567	24	ABK12893	Human proteinase PR
C	96	17	77.3	3642	24	ABK10756	DNA encoding human
C	97	17	77.3	3651	19	AAV42910	Human catechol-O-m
C	98	17	77.3	3651	24	ABN97375	Gene #1873 used to
C	99	17	77.3	3651	24	ABK10755	DNA encoding human
C	100	17	77.3	3660	24	ABK10754	DNA encoding human
C	101	17	77.3	3669	24	ABK10753	DNA encoding human
C	102	17	77.3	3672	24	ABK48893	DNA encoding human
C	103	17	77.3	3699	25	ABK10758	DNA encoding human
C	104	17	77.3	3699	25	AB276003	Human aggrucanase
C	105	17	77.3	3708	24	ABK10757	DNA encoding human
C	106	17	77.3	3750	24	ABK10760	DNA encoding human
C	107	17	77.3	3759	24	ABK10759	DNA encoding human
C	108	17	77.3	3846	24	ABK69988	CDNA encoding huma
C	109	17	77.3	4000	23	AB105442	Drosophila melanog
C	110	17	77.3	4080	25	AB276001	Human aggrucanase
C	111	17	77.3	4213	24	ABK37934	CDNA encoding huma
C	112	17	77.3	4277	24	ABK10761	Human proteinase pol
C	113	17	77.3	4380	23	AB122712	Drosophila melanog
C	114	17	77.3	4492	23	AB118340	Drosophila melanog
C	115	17	77.3	4589	23	AB105995	Drosophila melanog
C	116	17	77.3	5344	24	ABK87749	Human CDNA encodin
C	117	17	77.3	5511	24	ABK58371	Protein modificati
C	118	17	77.3	5701	24	ABK87748	Human CDNA encodin
C	119	17	77.3	5705	23	AB117266	Drosophila melanog
C	120	17	77.3	5710	23	AB118813	Drosophila melanog

ALIGNMENTS

RESULT 1

AAH03060/c
ID AAH03060 standard; DNA; 22 BP.

AAH03060;
15-JUN-2001 (first entry)

Microorganism detection method related oligonucleotide SEQ ID NO: 84.

Microorganism identification; pathogen; DNA sequencing; HLA type;
bi-directional sequencing; infection; mutation detection; PCR primer; ss.

Unidentified.

US614555.B1

10-APR-2001

13-MAY-1999; 99US-0311260.

01-MAY-1996; 96US-0640672.

19-JUL-1996; 96US-0684498.

22-FEB-1997; 97US-0807138.

20-JAN-1998; 98US-0009483.

(VISI-) VISIBLE GENETICS INC.

Leushner J, Hui M, Dunn JM, Lacroix J,

XX	DR	WPI, 2001-289716/30.
XX	XX	Composition for detecting microorganisms, comprising deoxynucleotide
XX	PT	triphosphates, dideoxynucleotide triphosphate, and thermostable
XX	PT	polymerase to incorporate dideoxynucleotide triphosphate into extending
XX	PT	polymerase
XX	PS	Disclosure; Column 63; 62pp; English.
XX	XX	The present invention provides a composition containing 4 dNTPs and at
XX	CC	least one dNTP and a thermally stable polymerase which incorporates
XX	CC	dNTPs into an extending nucleic acid polymer at a rate of not less than
XX	CC	0.4 times the rate of dNTP incorporation. This can be used with the PCR
XX	CC	primers provided in the invention to detect the presence of
XX	CC	microorganisms, such as Chlamydia trachomatis, HIV or human
XX	CC	papillomavirus, in a sample. In addition, it can be used to detect
XX	CC	mutations in a specific gene, to determine HLA type, and to produce
XX	CC	sequencing fragments for further study.
XX	SQ	Sequence 22 BP; 2 A; 3 C; 10 G; 7 T; 0 other;
XX	XX	Query Match 100.0%; Score 22; DB 22; Length 22;
XX	XX	Best Local Similarity 100.0%; Pred. No. 0.29;
XX	XX	Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	QY	1 ATTGCACACACACACACCA 22
XX	DB	22 ATTGCACACACACACACCA 1

RESULT 2

ABA00273
ID ABA00273 standard; DNA; 22 BP.

ABA00273;

29-NOV-2002 (first entry)

EBNA 2 antisense primer.

Primer; amplify; PCR; probe; detection; Epstein-Barr virus; EBV; ss.

Epstein-Barr virus.

WO200264842-A2.

22-AUG-2002.

13-FEB-2002; 2002WO-US04339.

13-FEB-2001; 2001US-268439P.

(CHIL-) CHILDRENS HOSPITAL RES FOUND.

Witte DP, Groen PA;

WPI, 2002-667015/71.

New compositions comprising nucleic acid sequences which specifically

hybridizes to Epstein-Barr virus (EBV) nucleic acid, for detecting EBV

in clinical specimens to determine patients at high risk of to

developing EBV infections

Claim 1, Page 44; 59pp; English.

The sequences given in ABA00268-75 are primers and probes which were

used in the compositions of the invention for the detection of

Epstein-Barr virus (EBV). The compositions comprise at least one

sequence which specifically hybridizes to a nucleic acid

sequence which complements and specifically hybridizes to EBV nucleic

acid. The oligonucleotide sequences and compositions comprising them

are useful for detecting EBV in clinical specimens to determine

CC patients who are at high risk to develop serious and costly medical
CC complications, and allow for better clinical management of these
CC patients by earlier recognition of their infection status. The
CC oligonucleotide sequences may also be used to amplify EBV DNA
CC sequences. The use of the oligonucleotide sequences in the assay for
CC detecting EBV has a broad dynamic range of detection from less than
CC 10-10000000 copies of EBV DNA, is less labour intensive, requiring only
CC one reaction tube for the EBV determination, highly sensitive, accurate
CC and has a rapid turn around time with assays that are completed,
CC including amplification, probe specific hybridization, and calculation
CC of copy number in less than 1 hour. The method may be adapted to
CC automated systems.

XX Sequence 22 BP; 7 A; 10 C; 3 G; 2 T; 0 other;

XX Query Match 100.0%; Score 22; DB 24; Length 22;

XX Best Local Similarity 100.0%; Pred. No. 0.29;

XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGCGCACCGACGACGACCA 22

DB 1 ATGCGCACCGACGACGACCA 22

RESULT 3

ABAS0785/c

ID ABAS0785 standard; DNA; 121 BP.

AC ABAS0785;

DT 01-FEB-2002 (first entry)

DE Human breast cell single exon nucleic acid probe #9480.

XX Human; microarray; single exon probe; gene expression; breast;

XX disease; cancer; ss.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,

XX useful for measuring gene expression in sample derived from human

XX breast, comprises number of single exon nucleic acid probes -

XX Claim 4, SEQ ID NO 9480; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon

XX nucleic acid probes for measuring gene expression in a sample derived

XX from human breast and BT 474 cells. The method involves contacting

XX the probes with a collection of detectably labelled nucleic acids

XX derived from mRNA of human breast, and then measuring the label

XX bound to each probe of the microarray. The probes are useful for

XX verifying the expression of regions of genomic DNA predicted to

XX encode proteins. They are useful for gene discovery, and for

CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIFO at ftp.wifo.int/pub/published_pct_sequences.

XX Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;

XX Query Match 86.4%; Score 19; DB 22; Length 121;

XX Best Local Similarity 100.0%; Pred. No. 6.6; Indels 0; Gaps 0;

XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GCCACGACGACGACGACCA 22

DB 111 GCCACGACGACGACGACCA 93

RESULT 4

ABA68749/c

ID ABA68749 standard; DNA; 121 BP.

AC ABA68749;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #17054.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-493447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

XX Claim 4, SEQ ID NO 17054; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX foetal liver. The present sequence is a single exon nucleic acid

XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIFO at ftp.wifo.int/pub/published_pct_sequences.

XX Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;

Query Match 86.4%; Score 19; DB 22; Length 121;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCCAGCAGCAGCA 22
 DB 111 GCCACCCAGCAGCAGCA 93

RESULT 5

ABAS715/c
 ID ABAS715 standard; DNA; 121 BP.

AC ABA35715;
 XX 23-JAN-2002 (first entry)

DE Probe #14181 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KM congenital heart disease; ss.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-486899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -

PS Claim 4; SEQ ID No 14181; 530bp; English.

XX The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart. The

CC present sequence is one such probe. The probes may be used for

CC predicting, measuring and displaying gene expression in samples derived

CC from the human heart via microarray. By measuring gene expression, the

CC probes are useful for predicting, diagnosing, grading, staging,

CC monitoring and prognosing diseases of the human heart and vascular system

CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

CC congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;

SO Query Match 86.4%; Score 19; DB 22; Length 121;

Best Local Similarity 100.0%; Pred. No. 6.6;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCCAGCAGCAGCA 22

DB 111 GCCACCCAGCAGCAGCA 93

RESULT 6
 AAK17094/c
 ID AAK17094 standard; DNA; 121 BP.

AC AAK17094;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 17085.

XX Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KM epilepsy; cancer; ss.

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

PT Single exon nucleic acid probes for analyzing gene expression in human

PT brains -

PS Example 4; SEQ ID NO: 17085; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,

CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

CC epilepsy and cancers. The present sequence is one of the probes of the

CC invention.

XX Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;

SO Query Match 86.4%; Score 19; DB 22; Length 121;

Best Local Similarity 100.0%; Pred. No. 6.6;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCCAGCAGCAGCA 22

DB 111 GCCACCCAGCAGCAGCA 93

RESULT 7
 AAK42877/c
 ID AAK42877 standard; DNA; 121 BP.

AC AAK42877;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 17434.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukemia; lymphoma; myeloma; ss.

```

XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 17434; 658bp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukemia and myeloma. The present sequence is one of
XX CC the probes of the invention.
XX SQ Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;

Query Match      86.4%; Score 19; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCA 22
   |||||
Db 111 GCCACCACGACGACCA 93

RESULT 8
AA123642/C
ID AA123642 standard; DNA; 121 BP.
XX AC AA123642;
XX DT 12-OCT-2001 (first entry)
XX DE Probe #13575 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00670.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000US-0236359.

```

```

PR 04-OCT-2000; 2000GB-0024263.
XX OS (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488901/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX PS Claim 25; SEQ ID No 13575; 487bp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENPs). The present sequence is one such probe. The SENPs are derived
XX CC from human Hela cells. The SENPs can be used to produce a single exon
XX CC microarray, which can be used for measuring human gene expression in a
XX CC sample derived from human cervical epithelial cells. By measuring gene
XX CC expression, the probes are therefore useful in grading and/or staging
XX CC of diseases of the cervix, notably cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;

Query Match      86.4%; Score 19; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCA 22
   |||||
Db 111 GCCACCACGACGACCA 93

RESULT 9
AA148952/C
ID AA148952 standard; DNA; 121 BP.
XX AC AA148952;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #17638 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-48897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -

```

PS Claim 25; SEQ ID No 17638; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;
Query Match 86.4%; Score 19; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
4 GCCACGACGACGACGACCA 22
111 GCCACGACGACGACGACCA 93
Db
RESULT 10
AA109257/c
ID AA109257 standard; DNA; 121 BP.
XX
AC AA109257;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #9248 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
FT Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX Claim 25; SEQ ID No 9248; 322bp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer; disorders of development, inflammatory diseases
CC of the breast; fibrocystic changes; proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WFO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;
Query Match 86.4%; Score 19; DB 22; Length 121;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
4 GCCACGACGACGACGACCA 22
111 GCCACGACGACGACGACCA 93
Db
RESULT 11
ABS42507/c
ID ABS42507 standard; DNA; 121 BP.
XX
AC ABS42507;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID No 17497.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00664.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488698/53.
XX
FT Human genome-derived single exon nucleic acid probes useful for
XX analysing gene expression in human adult liver -
XX
XX Claim 4; SEQ ID No 17497; 658bp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult
CC liver. (I) may be used for predicting, measuring and displaying gene
CC expression in samples derived from human adult liver. The genes
CC identified may be involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
CC is associated with coronary heart disease. ABS25011-ABS51005 represent
CC human liver single exon nucleic acid probes of the invention.
CC Note: The sequence information for this patent does not appear in the
CC printed specification but was obtained in electronic format directly
CC from WFO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;
Query Match 86.4%; Score 19; DB 23; Length 121;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGCACCCAGCAGCACCA 22
 |||||
 Db 111 GGCACCCAGCAGCACCA 93

RESULT 12
 ABS16932/c
 ABS16932 standard; DNA; 121 BP.

AC ABS16932;
 XX
 XX 19-AUG-2002 (first entry)
 XX Human genome-derived single exon probe ORF from lung SEQ ID No 16923.
 DE
 XX
 XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.

XX Homo sapiens.
 OS
 XX
 XX W0200186003-A2.
 PN
 XX
 XX 15-NOV-2001.
 PD
 XX
 XX 30-JAN-2001; 2001MO-US00665.
 PF
 XX
 XX 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2002-114183/15.
 DR
 XX
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 PT
 XX
 XX Claim 4; SEQ ID No 16923; 634bp; English.

CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIP0 at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX
 XX Sequence 121 BP; 26 A; 17 C; 46 G; 32 T; 0 other;
 SQ

Query Match 86.4%; Score 19; DB 24; Length 121;
 Best Local Similarity 100.0%; Pred.No. 6.6; Mismatches 0;
 Matches 19; Conservative 0; Indels 0; Gaps 0;

QY 4 GGCACCCAGCAGCACCA 22
 |||||
 Db 111 GGCACCCAGCAGCACCA 93

RESULT 13
 ABA45663/c
 ID ABA45663 standard; DNA; 386 BP.
 XX
 XX ABA45663;
 AC
 XX
 XX 01-FEB-2002 (first entry)
 DT
 XX
 XX Human breast cell single exon nucleic acid probe #4358.
 DE
 XX
 XX Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 KW
 XX
 XX Homo sapiens.
 OS
 XX
 XX W0200157271-A2.
 PN
 XX
 XX 09-AUG-2001.
 PD
 XX
 XX 30-JAN-2001; 2001MO-US00662.
 PF
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-496933/54.
 DR
 XX
 XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 PT
 XX
 XX Claim 1; SEQ ID NO 4358; 327bp + sequence listing; English.
 PS
 XX The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 386 BP; 91 A; 87 C; 84 G; 124 T; 0 other;

Query Match 86.4%; Score 19; DB 22; Length 386;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCA 22
DB 350 GCCACCACGACGACCA 332

RESULT 14
ID ABA56163/C
ID ABA56163 standard; DNA; 386 BP.

AC ABA56163; 2

DT 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #4468.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

XX WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Claim 1; SEQ ID NO 4468; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX foetal liver. The present sequence is a single exon nucleic acid

CC Probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 386 BP; 91 A; 87 C; 84 G; 124 T; 0 other;

Query Match 86.4%; Score 19; DB 22; Length 386;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCA 22
DB 350 GCCACCACGACGACCA 332

RESULT 15
ID ABA25816/C
ID ABA25816 standard; DNA; 386 BP.

AC ABA25816;

DT 23-JAN-2002 (first entry)

XX Probe #4282 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

XX cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

OS Homo sapiens.

XX WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488699/53.

XX Claim 1; SEQ ID No 4282; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for

XX measuring human gene expression in a sample derived from human heart. The

XX present sequence is one such probe. The probes may be used for

XX predicting, measuring and displaying gene expression in samples derived

XX from the human heart via microarrays. By measuring gene expression, the

XX probes are useful for predicting, diagnosing, grading, staging,

XX monitoring and prognosing diseases of the human heart and vascular system

XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

XX congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 386 BP; 91 A; 87 C; 84 G; 124 T; 0 other;

XX Query Match 86.4%; Score 19; DB 22; Length 386;

Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCA 22
DB 350 GCCACCACGACGACCA 332

RESULT 16

AAK04355/C
ID AAK04355 standard; DNA; 386 BP.

AC AAK04355;

DT 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 4346.

XX Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX epilepsy; cancer; ss.

OS Homo sapiens.

PN WO200157275-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

PT brains -

XX Example 4; SEQ ID NO: 4346; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,

CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

CC epilepsy and cancers. The present sequence is one of the probes of the

CC invention.

XX Sequence 386 BP; 91 A; 87 C; 84 G; 124 T; 0 other;

Query Match 86.4%; Score 19; DB 22; Length 386;

Best Local Similarity 100.0%; Pred. No. 6.6;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCA 22
DB 350 GCCACCACGACGACCA 332

RESULT 17

AAK29849/C
ID AAK29849 standard; DNA; 386 BP.

XX AAK29849;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 4406.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

PN WO200157276-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 4406; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukaemia and myeloma. The present sequence is one of

CC the probes of the invention.

XX Sequence 386 BP; 91 A; 87 C; 84 G; 124 T; 0 other;

Query Match 86.4%; Score 19; DB 22; Length 386;

Best Local Similarity 100.0%; Pred. No. 6.6;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCA 22
DB 350 GCCACCACGACGACCA 332

XX RESULT 18

AAI1440/C

ID AAI14440 standard; DNA; 386 BP.

XX AAI14440;

XX 12-OCT-2001 (first entry)

XX Probe #4373 for gene expression analysis in human cervical cell sample.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488901/53.
 DR WPI; 2001-488901/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX Claim 25; SEQ ID No 4373; 487bp; English.
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human HeLa cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 CC
 XX Sequence 386 BP; 91 A; 87 C; 84 G; 124 T; 0 other;
 SQ
 Query Match 86.4%; Score 19; DB 22; Length 386;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 GCCACCAACGACGACCA 22
 Db 350 GCCACCAACGACGACCA 332
 RESULT 19
 AA135812/C
 ID AA135812 standard; DNA; 386 BP.
 AC AA135812;
 XX
 AC AA135812;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #4498 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 OS WO200157272-A2.
 XX
 EN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488901/53.
 DR WPI; 2001-488901/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX Claim 25; SEQ ID No 4498; 654bp; English.
 CC The present invention relates to single exon nucleic acid probes (SENPs).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 CC
 XX Sequence 386 BP; 91 A; 87 C; 84 G; 124 T; 0 other;
 SQ
 Query Match 86.4%; Score 19; DB 22; Length 386;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 GCCACCAACGACGACCA 22
 Db 350 GCCACCAACGACGACCA 332
 RESULT 20
 AA104263/C
 ID AA104263 standard; DNA; 386 BP.
 AC AA104263;
 XX
 AC AA104263;
 XX
 DT 09-OCT-2001 (first entry)
 XX
 DE Probe #4254 used to measure gene expression in human breast sample.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 OS WO200157270-A2.
 XX
 EN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US00661.
 XX
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-476286/51.
 DR WPI; 2001-476286/51.
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 25; SEQ ID No 4254; 322bp; English.
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases

CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocytic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 386 BP; 91 A; 87 C; 84 G; 124 T; 0 other;

Query Match 86.4%; Score 19; DB 22; Length 386;

Best Local Similarity 100.0%; Pred. No. 6.6;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCA 22

DB 350 GCCACCACGACGACCA 332

RESULT 21

ABS29496/C

ID ABS29496 standard; DNA; 386 BP.

XX ABS29496;

DT 25-FEB-2003 (first entry)

XX Human liver single exon probe, SEQ ID No 4486.

DE Human liver single exon probe, SEQ ID No 4486.

XX Human; single exon nucleic acid probe; liver; cirrhosis;

KW hyperlipoproteinemia; hyperlipidaemia; hypercholesterolemia;

KM coronary heart disease; ss.

XX Homo sapiens.

OS WO200157273-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00664.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488998/53.

XX Claim 1; SEQ ID No 4486; 658bp; English.

CC The invention relates to a single exon nucleic acid probe (SRNP) (1) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13103 defined nucleotide sequences given in the
CC specification (or complements/fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult
CC liver. (1) may be used for predicting, measuring and displaying gene
CC expression in samples derived from human adult liver. The genes
CC identified may be involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinemia, hyperlipidaemia and hypercholesterolemia which
CC is associated with coronary heart disease. ABS25011-ABS51005 represent
CC human liver single exon nucleic acid probes of the invention.
CC Note: The sequence information for this patent does not appear in the
CC printed specification but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 386 BP; 91 A; 87 C; 84 G; 124 T; 0 other;

Query Match 86.4%; Score 19; DB 22; Length 386;

Best Local Similarity 100.0%; Pred. No. 6.6;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCA 22

DB 350 GCCACCACGACGACCA 332

RESULT 22

ABS04410/C

ID ABS04410 standard; DNA; 386 BP.

XX ABS04410;

DT 19-AUG-2002 (first entry)

XX Human genome-derived single exon probe from lung SEQ ID No 4401.

DE Human; ds; single exon probe; aetna; lung cancer; COPD; ILB;

KW chronic obstructive pulmonary disease; interstitial lung disease;

KM familial idiopathic pulmonary fibrosis; neurofibromatosis;

KM tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karsenger syndrome;

KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KM primary ciliary dyskinesia; pulmonary hypertension;

KM hyaline membrane disease.

XX Homo sapiens.

OS WO200186003-A2.

PN 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US00665.

XX 04-FEB-2000; 2000US-180312P.

XX 26-MAY-2000; 2000US-207456P.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

XX Claim 1; SEQ ID No 4401; 634bp; English.

CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12614 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes. The novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably

CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
 CC polymerase chain reaction, and for production of transgenic plants. (1),
 CC or antibodies that recognize their expression products, are used for
 CC detecting the presence of KT2440, particularly in presence of other,
 CC even closely related, bacteria. KT2440 is one of the bacteria classified
 CC as safe, by the National Institutes of Health, for genetic engineering
 CC work, e.g. as microbial production strains, for biological remediation
 CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant
 CC homology with sequences in other bacteria (specifically the closely
 CC related pathogen *P. aeruginosa*). Compared with other 'safe' bacteria, it
 CC has greater catabolic activity and better survival in, and adaptation to,
 CC the rhizosphere and soil.

SO Sequence 2001 BP; 333 A; 719 C; 619 G; 325 T; 5 other;
 Query Match 86.4%; Score 19; DB 22; Length 2001;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCAACGACGACCA 22
 DB 1001 GCCACCAACGACGACCA 1019

RESULT 25
 AAF61083/C
 ID AAF61083 standard; DNA; 2028 BP.
 XX AAF61083;
 AC AAF61083;
 XX 16-MAY-2001 (first entry)
 DT
 XX
 DE *P. putida* KT2440-associated DNA CEF09883.
 XX
 XX Transgenic plant; detection; probe; amplification; vaccine carrier;
 KM microbial production strain; biological remediation; ds.
 XX
 OS *Pseudomonas putida*.
 XX
 XX DE19935088-A1.
 PN
 XX 01-FEB-2001.
 PD
 XX 27-JUL-1999; 99DE-1035088.
 PF
 XX 27-JUL-1999; 99DE-1035088.
 PR
 XX 27-JUL-1999; 99DE-1035088.
 XX
 PA (TIGR-) TIGR INST GENOMIC RES.
 PA (QUTR-) QUTR GENOME.
 PA (GBF-) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.
 PA (DKF-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
 PA (MED-) MEDIZINISCHE HOCHSCHULE HANNOVER.
 XX
 DR WPI; 2001-192469/20.
 XX
 PT New DNA sequences specific for *Pseudomonas putida* KT2440, useful as
 PT safe genetic engineering host, allow detection in presence of other
 PT related bacteria -
 XX
 PS Claim 1a; Page 135-136; 158pp; German.
 XX
 CC This invention describes novel DNA sequences (1) for specific detection
 CC of *Pseudomonas putida* KT2440. The invention also describes (1)
 CC recombinant expression vector containing (1) (2) prokaryotic or
 CC eukaryotic cells transformed or transfected with (1) or the vector of
 CC (1); (3) production of expression products by culturing cells of (2);
 CC (4) expression products, or their fragments, of (1) and synthetic
 CC proteins or peptides with the same sequences (A); (5) poly- or
 CC monoclonal antibodies (Ab) that react specifically with (A); (6)
 CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
 CC plants that contain transformed or transfected cells of (2); (8)
 CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips
 CC carrying one or more (1), (1), and their fragments, are used as probes

CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
 CC polymerase chain reaction, and for production of transgenic plants. (1),
 CC or antibodies that recognize their expression products, are used for
 CC detecting the presence of KT2440, particularly in presence of other,
 CC even closely related, bacteria. KT2440 is one of the bacteria classified
 CC as safe, by the National Institutes of Health, for genetic engineering
 CC work, e.g. as microbial production strains, for biological remediation
 CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant
 CC homology with sequences in other bacteria (specifically the closely
 CC related pathogen *P. aeruginosa*). Compared with other 'safe' bacteria, it
 CC has greater catabolic activity and better survival in, and adaptation to,
 CC the rhizosphere and soil.

SO Sequence 2028 BP; 331 A; 628 C; 710 G; 356 T; 3 other;
 Query Match 86.4%; Score 19; DB 22; Length 2028;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCAACGACGACCA 22
 DB 869 GCCACCAACGACGACCA 851

RESULT 26
 ABL16704
 ID ABL16704 standard; DNA; 2889 BP.
 XX ABL16704;
 AC ABL16704;
 XX 26-MAR-2002 (first entry)
 DT
 XX
 DE *Drosophila melanogaster* genomic polynucleotide SEQ ID NO 1585.
 XX
 XX *Drosophila*; developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ds.
 XX
 OS *Drosophila melanogaster*.
 XX
 XX WC200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001MO-US09231.
 PF
 XX 23-MAR-2001; 2000US-191637P.
 PR
 XX 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PT WPI; 2001-656860/75.
 DR
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 1585; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABBS7737-ABBS72072).
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 2889 BP; 863 A; 695 C; 561 G; 770 T; 0 other;

Query Match 86.4%; Score 19; DB 23; Length 2889;
 Best Local Similarity 100.0%; Pred. No. 6.6;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GCCACACACGACGACCA 22
 DB 1034 GCCACACACGACGACCA 1052

RESULT 27
 ID AAF75404/C
 XX AAF75404 standard; DNA; 123 BP.

AC AAF75404;
 XX AAF75404;
 DT 14-MAY-2001 (first entry)

DE Codon-optimised HPV16 L1 fragment MN4A14.

KM Human papillomavirus; HPV, HPV16; HPV6a; HPV18; L1; E2; E7; E1;
 KM antiviral; immunostimulant; vaccine; immunogen; infection; ss.

OS Human papillomavirus.
 OS Synthetic.

PN WO200114416-A2.

XX 01-MAR-2001.

PF 21-AUG-2000; 2000WO-US22932.

PR 25-AUG-1999; 99US-0150728.
 PR 07-JUN-2000; 2000US-0210143.

XX (MERI) MERCK & CO INC.

XX Neoper MP, McClements WL, Jansen KU, Schultz LD, Chen L, Wang X;
 PI WPI; 2001-218428/22.

PT Novel synthetic polynucleotide encoding human papillomavirus (HPV)
 PT protein or mutated HPV protein useful as anti-HPV vaccines, comprises
 PT optimized-codons for expression of the viral proteins in human host
 cells

XX Example 2; Fig 17; 119pp; English.

CC The present sequence is an oligomer which was used in the assembly of
 CC one of a number of synthetic polynucleotides that encode a human
 CC papillomavirus (HPV) protein, or a mutated form of a HPV protein. The
 CC mutated HPV proteins have reduced protein function as compared to wild
 CC type proteins but maintain immunogenicity. The proteins comprise codons
 CC for optimised expression in humans. The polynucleotides are useful as a
 CC vaccine which provides effective immunoprophylaxis against
 CC papillomavirus infection through stimulation of neutralising antibody
 CC and cell-mediated immunity.

XX Sequence 123 BP; 12 A; 30 C; 46 G; 35 T; 0 other;

Query Match 81.8%; Score 18; DB 22; Length 123;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CCACGACGACGACGACCA 22
 DB 71 CCACGACGACGACGACCA 54

RESULT 28
 ABK14577
 ID ABK14577 standard; DNA; 180 BP.
 XX

AC ABK14577;

XX 08-MAY-2002 (first entry)

DE Human papilloma virus, HPV-16, synthetic L1 gene oligonucleotide #9.

KM HPV-16; L1; late gene; mutant; ss; DNA vaccine; virucide; genital wart;
 KM codon optimisation; anogenital condyloma; squamous intraepithelial lesion;
 KM cervical cancer; upper aerodigestive tract carcinoma; penile carcinoma;
 KM vulval carcinoma; anal carcinoma; mutant.

OS Human papillomavirus.
 OS Synthetic.

PN WO200202142-A1.

PD 10-JAN-2002.

PF 29-JUN-2001; 2001WO-SE01501.

PR 03-JUL-2000; 2000SE-0002498.

XX (SCHW/) SCHWARTZ S.

PA Schwartz S;

DR WPI; 2002-164498/21.

PT Novel human papilloma virus vaccine useful to protect humans against
 PT HPV infection, has HPV nucleic acid encoding HPV protein expressed in
 PT human cells leading to immune response

XX Example; Fig 2; 28pp; English.

CC A human papilloma virus (HPV) vaccine comprising a HPV
 CC nucleic acid which encodes HPV protein which is expressed in human
 CC cells and which leads to an immune response. The HPV nucleic acid
 CC has its sequence altered to remove negative regulatory elements but
 CC encodes the same amino acid sequence. The nucleic acid is cloned
 CC into a human expression vector including elements necessary for
 CC expressing the HPV sequence. The sequence may be from HPV 6, 11, 16, 18,
 CC 31, 33 and/or 45. The vaccine is useful against HPV infections e.g.
 CC genital warts, anogenital condyloma, squamous intraepithelial lesion,
 CC cervical cancer, upper aerodigestive tract carcinoma, carcinomas of
 CC the penis, vulva or anus. The present sequence is a synthetic
 CC oligonucleotide representing a part of a HPV-16 late gene 1 (L1) whose
 CC codons have been optimised to remove the negative regulatory elements.
 CC PCR reactions are then used to assemble the complete synthetic HPV-16 L1
 CC gene.

XX Sequence 180 BP; 45 A; 67 C; 48 G; 20 T; 0 other;

Query Match 81.8%; Score 18; DB 24; Length 180;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CCACGACGACGACGACCA 22
 DB 122 CCACGACGACGACGACCA 139

RESULT 29
 ABL29435
 ID ABL29435 standard; DNA; 1059 BP.

XX ABL29435;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 39778.
 XX Drosophila developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; de.

```

XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PA 11-JUL-2000; 2000US-0614150.
XX PE (PEXE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PMD, Myers EW;
XX WI WPI; 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1, SEQ ID NO 38778; 21pp + Sequence listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX CC sequences (AB161840-AB16175) and the encoded proteins
XX CC (AB161737-AB1617072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1059 BP; 247 A; 280 C; 299 G; 233 T; 0 other;

Query Match      81.8%; Score 18; DB 23; Length 1059;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 CGACCCAGCAGCAGCACA 22
Db      1007 CCACCCAGCAGCAGCACA 1024

RESULT 30
AAH73845
ID AAH73845 standard; DNA; 1158 BP.
XX AC AAH73845;
XX DT 28-SEP-2001 (first entry)
XX DE Permuterin protein coding sequence #6.
XX KM Insecticidal; immunosuppressive; potato; patatin; enzyme;
XX KM allergenicity; larva growth; lipid acyl hydrolase; insecticide; ds.
XX OS Unidentified.
XX OS WO200149834-A2.
XX FN 12-JUL-2001.
XX PD 05-JAN-2001; 2001WO-US00342.
XX PF 06-JAN-2000; 2000US-0174669.
XX PR (MONS ) MONSANTO CO.
XX PA Alibhai MF, Astwood JD, Mcwhorter CA, Sampson HA;
XX PI

```

```

XX DR WPI; 2001-441874/47.
XX PT Modified potato patatin proteins with reduced antigenicity, useful as
XX PT insecticides for controlling e.g. round worm and root worm -
XX PS Example 14; Pages 194-195; 223pp; English.
XX CC The present invention relates to modified potato patatins that maintain
XX CC enzymatic and insecticidal activity but which have reduced allergenicity.
XX CC Groups (especially Tyr) which bind to anti-patatin antibodies were
XX CC identified and glycosylation sites involved in antibody binding were
XX CC removed via site directed mutagenesis. The patatins stunt the growth of
XX CC larvae so that maturation is prevented or delayed. The patatins also have
XX CC non-specific lipid acyl hydrolase activity. The modified patatins are
XX CC also useful for inhibiting the activity of corn round worms.
XX CC Deallergised protein can be used as insecticides, as nutritional
XX CC supplements and as immunising agents. The present sequence was used to
XX CC illustrate the present invention.
XX SQ Sequence 1158 BP; 260 A; 400 C; 293 G; 205 T; 0 other;

Query Match      81.8%; Score 18; DB 22; Length 1158;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 GCCACCCAGCAGCAGCACC 21
Db      46 GCCACCCAGCAGCAGCACC 63

RESULT 31
AAQ72603
ID AAQ72603 standard; cDNA; 1164 BP.
XX AC AAQ72603;
XX DT 25-MAR-2003 (updated)
XX DT 18-MAY-1995 (first entry)
XX DE Plant insecticidal patatin structural cDNA.
XX KM Insecticidal patatin; genetic transformation of plants;
XX KM European corn borer; western and southern corn rootworm; boll weevil;
XX KM Colorado potato beetle; ss.
XX OS Solanum cardiophyllum.
XX OS WO9421805-A2.
XX FN 29-SEP-1994.
XX PD 02-MAR-1994; 94WO-US02306.
XX PF 12-MAR-1993; 93US-0031146.
XX PR (MONS ) MONSANTO CO.
XX PA Brown SM, Greenplate JT, Isaac BG, Jennings MG;
XX PI Levine EB, Purcell JP;
XX WI WPI; 1994-317025/39.
XX PT Controlling plant-eating insect infestations - by providing an
XX PT insecticidal patatin for ingestion by the insect opt. by
XX PT genetically transforming plants
XX PS Claim 8; Page 52; 52pp; English.
XX CC AAQ72603 is a patatin structural cDNA, it was used in the construction
XX CC of a recombinant ds cDNA, which contained a plant cell RNA promoter,
XX CC a patatin structural coding sequence and a 3'UTR. This cDNA is
XX CC specifically designed to genetically transform corn and other monocots

```

CC enabling them to produce insecticidal patatin. The patatin can be used
CC to control various plant eating insects including western and southern
CC corn rootworm, boll weevil, Colorado potato beetle and European corn
CC borer. It is also lethal to some larvae and will stunt the growth of
CC survivors preventing or severely delaying metamorphosis, therefore
CC preventing reproduction.
CC (Updated on 25-MAR-2003 to correct FN field.)

XX
SQ Sequence 1164 BP; 263 A; 400 C; 294 G; 207 T; 0 other;

Query Match 81.8%; Score 18; DB 15; Length 1164;
Best Local Similarity 100.0%; Pred. No. 19;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCC 21
|||

DB 46 GCCACCACGACGACCC 63
|||

RESULT 32

ID AAH73850 standard; DNA; 1167 BP.

XX AC AAH73850;

XX DT 28-SEP-2001 (first entry)

XX DE Permutin protein coding sequence #7.

XX KM Insecticidal; immunosuppressive; potato; patatin; enzyme;
XX KM allergenicity; larva growth; lipid acyl hydrolase; insecticide; ds.

XX OS Unidentified.

XX PN WO200149834-A2.

XX PD 12-JUL-2001.

XX PF 05-JAN-2001; 2001WO-US00342.

XX PR 06-JAN-2000; 2000US-0174669.

XX PA (MONS) MONSANTO CO.

XX PI Alibhai MF, Astwood JD, McWherter CA, Sampson HA;

XX DR WPI; 2001-441874/47.

XX PT Modified potato patatin proteins with reduced antigenicity, useful as
XX PT insecticides for controlling e.g. round worm and root worm -

XX PS Example 16; Pages 197-198; 223pp; English.

XX CC The present invention relates to modified potato patatins that maintain
XX CC enzymatic and insecticidal activity but which have reduced allergenicity.
XX CC Groups (especially Tyr) which bind to anti-patatin antibodies were
XX CC identified and glycosylation sites involved in antibody binding were
XX CC removed via site directed mutagenesis. The patatins stunt the growth of
XX CC larvae so that maturation is prevented or delayed. The patatins also have
XX CC non-specific lipid acyl hydrolase activity. The modified patatins are
XX CC also useful for inhibiting the activity of corn round worms.
XX CC Deallergens protein can be used as insecticides, as nutritional
XX CC supplements and as immunising agents. The present sequence was used to
XX CC illustrate the present invention.

XX SQ Sequence 1167 BP; 263 A; 402 C; 297 G; 205 T; 0 other;

Query Match 81.8%; Score 18; DB 22; Length 1167;
Best Local Similarity 100.0%; Pred. No. 19;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCC 21
|||

DB 46 GCCACCACGACGACCC 63
|||

RESULT 33

ID AAH73853 standard; DNA; 1167 BP.

XX AC AAH73853;

XX DT 28-SEP-2001 (first entry)

XX DE Permutin protein coding sequence #8.

XX KM Insecticidal; immunosuppressive; potato; patatin; enzyme;
XX KM allergenicity; larva growth; lipid acyl hydrolase; insecticide; ds.

XX OS Unidentified.

XX PN WO200149834-A2.

XX PD 12-JUL-2001.

XX PF 05-JAN-2001; 2001WO-US00342.

XX PR 06-JAN-2000; 2000US-0174669.

XX PA (MONS) MONSANTO CO.

XX PI Alibhai MF, Astwood JD, McWherter CA, Sampson HA;

XX DR WPI; 2001-441874/47.

XX PT Modified potato patatin proteins with reduced antigenicity, useful as
XX PT insecticides for controlling e.g. round worm and root worm -

XX PS Claim 3; Pages 200-201; 223pp; English.

XX CC The present invention relates to modified potato patatins that maintain
XX CC enzymatic and insecticidal activity but which have reduced allergenicity.
XX CC Groups (especially Tyr) which bind to anti-patatin antibodies were
XX CC identified and glycosylation sites involved in antibody binding were
XX CC removed via site directed mutagenesis. The patatins stunt the growth of
XX CC larvae so that maturation is prevented or delayed. The patatins also have
XX CC non-specific lipid acyl hydrolase activity. The modified patatins are
XX CC also useful for inhibiting the activity of corn round worms.
XX CC Deallergens protein can be used as insecticides, as nutritional
XX CC supplements and as immunising agents. The present sequence was used to
XX CC illustrate the present invention.

XX SQ Sequence 1167 BP; 263 A; 402 C; 297 G; 205 T; 0 other;

Query Match 81.8%; Score 18; DB 22; Length 1167;
Best Local Similarity 100.0%; Pred. No. 19;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACCC 21
|||

DB 46 GCCACCACGACGACCC 63
|||

RESULT 34

ID AAF75383 standard; DNA; 1518 BP.

XX AC AAF75383;

XX DT 14-MAY-2001 (first entry)

XX DE Codon-optimised HPV16 L1 gene.

XX KM Human papillomavirus; HPV; HPV16; HPV6a; HPV18; L1; E2; E7;
XX KM antiviral; immunostimulant; vaccine; immunogen; infection; ds.

OS Human papillomavirus.
 OS Synthetic.
 XX WO200114416-A2.
 XX PD 01-MAR-2001.
 XX PF 21-AUG-2000; 2000WO-US22932.
 XX PR 25-AUG-1999; 99US-0150728.
 XX PR 07-JUN-2000; 2000US-0210143.
 XX PA (MER1) MERCK & CO INC.
 XX PI Nepper MP, McClements WL, Jansen KU, Schultz LD, Chen L, Wang X;
 XX WPI; 2001-218428/22.
 XX DR
 XX PT Novel synthetic polynucleotide encoding human papillomavirus (HPV)
 XX protein or mutated HPV protein useful as anti-HPV vaccines, comprises
 XX optimized-codons for expression of the viral proteins in human host
 XX cells -
 XX PS Claim 7; Fig 1; 11pp; English.
 XX SQ
 CC The present sequence is one of a number of synthetic polynucleotides
 CC that encode a human papillomavirus (HPV) protein, or a mutated form of
 CC a HPV protein. The mutated HPV proteins have reduced protein function
 CC as compared to wild type proteins but maintain immunogenicity. The
 CC proteins comprise codons for optimised expression in humans. The
 CC polynucleotides are useful as a vaccine which provides effective
 CC immunoprophylaxis against papillomavirus infection through stimulation
 CC of neutralising antibody and cell-mediated immunity.
 XX SQ
 SQ Sequence 1518 BP; 332A; 566 C; 414 G; 206 T; 0 other;
 Query Match 81.8%; Score 18; DB 22; Length 1518;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 CCACCACGACGACGACCA 22
 Db 1466 CCACCACGACGACGACCA 1483
 RESULT 35
 ABL58977
 ID ABL58977 standard; DNA; 1518 BP.
 XX
 AC ABL58977;
 XX
 DT 22-JUL-2002 (first entry)
 XX
 DE HPV16-L1 encoding DNA 2.
 XX
 KM Human Papilloma virus; HPV; HPV16-L1; HPV16-L2; vaccine; cervical cancer;
 KM HPV16-L1; cytostatic; virucide; gene; ds.
 XX
 OS Human papillomavirus.
 XX
 FH Key Location/Qualifiers
 FT 1..1515
 FT /*tag= a
 FT /product= "HPV16-L1"
 XX
 EN WO200238769-A2.
 XX
 PD 16-MAY-2002.
 XX
 PF 19-SEP-2001; 2001WO-DE03618.
 XX
 PR 09-NOV-2000; 2000DE-1055545.
 XX

PA (DEK-) DEUT KREBSFORSCHUNGSZENTRUM
 PA (IPK-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
 XX
 XX PI Mueller M, Leder C, Kleinschmidt J, Sonnwald U, Biemelt S;
 XX WPI; 2002-426950/45.
 XX DR P-PSDB; ABB77479.
 XX
 XX PT New DNA sequences encoding human papilloma virus L1 or L2 protein,
 XX useful in vaccines, are optimized for high-level expression in
 XX eukaryotic cells -
 XX PS Claim 1; Fig 6; 39pp; German.
 XX SQ
 CC The invention relates to DNA sequences (I) that encode human papilloma
 CC virus (HPV) L1 protein (ABL58976, ABL58977) or L2 protein (ABL58978)
 CC or proteins with the biological activity of L1 and L2. Expression vectors
 CC containing (I) or a similar sequence encoding an L1/L2 fusion protein
 CC (ABL58979-ABL58981) and the proteins encoded by them (ABB77478-ABB77483),
 CC are useful in vaccines, especially for control of cervical cancer. (I)
 CC are also useful for recombinant production of L1 and L2 proteins. (I) are
 CC optimised for codon usage in eukaryotic cells and provide high yields of
 CC L1/L2 or their fusions, without the use of viral vectors.
 XX SQ
 SQ Sequence 1518 BP; 348 A; 538 C; 435 G; 197 T; 0 other;
 Query Match 81.8%; Score 18; DB 24; Length 1518;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 CCACCACGACGACGACCA 22
 Db 1466 CCACCACGACGACGACCA 1483

Search completed: August 14, 2003, 21:41:30
 Job time : 141.6 secs

Mon Aug 18 10:30:09 2003

us-10-074-620-5.011.rng

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 20:57:44 ; Search time 126 Seconds
(without alignments)
428,452 Million cell updates/sec

Title: US-10-074-620-5
Perfect score: 20
Sequence: 1 agggatgcctgcacacaga 20

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database :

N_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	24	ABN00272
2	18	90.0	20	22	AAH03059
3	17	85.0	20	13	AAQ02134
4	16	80.0	652	22	AAK68955
5	16	80.0	1452	25	ABK59264
6	16	80.0	1452	25	ABK59269
7	16	80.0	1458	25	ABK59270
8	16	80.0	1545	21	AAK60576

9	16	80.0	1545	21	AAK60577	Bacillus amylioliqu
10	16	80.0	1545	21	AAK60578	Bacillus amylioliqu
11	16	80.0	1545	21	AAK60579	Bacillus amylioliqu
12	16	80.0	1972	12	AAQ01032	EcORI-BamHI fragme
13	16	80.0	2083	18	AAV02472	Bacillus amylioliqu
14	16	80.0	2084	16	AAQ08067	Bacillus amylioliqu
15	16	80.0	2084	16	AAQ095032	Bacillus amylioliqu
16	16	80.0	2084	20	AAK59681	DNA encoding a ter
17	16	80.0	2084	20	AAK59681	Wild type Teramyl
18	16	80.0	2084	21	AAK48484	Bacillus amylioliqu
19	16	80.0	2084	24	ABL96211	Teramyl-like-alpha
20	16	80.0	2084	24	ABL50568	B. amylioliquetac
21	16	80.0	2084	24	AAI72215	Bacillus alpha amy
22	16	80.0	2084	24	AAK30026	Bacillus DNA encod
23	16	80.0	2604	20	AAK21079	Bacillus amylioliqu
24	16	80.0	2604	22	AAK37850	B. amylioliquetac
25	15	75.0	364	22	ABA47347	Human breast cell
26	15	75.0	364	22	ABA65232	Human foetal liver
27	15	75.0	364	22	ABA32333	Probe #10799 for g
28	15	75.0	364	22	AAK13649	Human brain expres
29	15	75.0	364	22	AAK39390	Human bone marrow
30	15	75.0	364	22	AAI20203	Probe #10136 for g
31	15	75.0	364	22	AAI45403	Probe #14089 used
32	15	75.0	364	22	AAI05907	Probe #5898 used
33	15	75.0	364	22	ABK38978	Human liver single
34	15	75.0	364	24	ABK13477	Human genome-deriv
35	15	75.0	393	22	ABA36167	Probe #14633 for g
36	15	75.0	416	21	AAK18000	Lung cancer associ
37	15	75.0	446	22	ABA42197	Human breast cell
38	15	75.0	446	22	ABA52619	Human foetal liver
39	15	75.0	446	22	ABA42407	Probe #873 for gen
40	15	75.0	446	22	ABA26282	Probe #4748 for ge
41	15	75.0	446	22	AAK00880	Human brain expres
42	15	75.0	446	22	AAK26335	Human bone marrow
43	15	75.0	446	22	AAI10968	Probe #901 for gen
44	15	75.0	446	22	AAI12228	Probe #914 used to
45	15	75.0	446	22	AAI00889	Probe #880 used to
46	15	75.0	446	23	ABK52526	Human liver single
47	15	75.0	446	23	ABK50923	Human genome-deriv
48	15	75.0	499	23	ABK48298	Human prostate exp
49	15	75.0	954	24	AAK48298	DNA encoding novel
50	15	75.0	1102	24	AAK29668	Human G-protein co
51	15	75.0	1207	23	AAK80857	DNA EST-derived
52	15	75.0	1293	22	AAK88420	Human EST-derived
53	15	75.0	1410	24	ABK84622	Human CDNA differe
54	15	75.0	1410	24	ABK66964	Lung cancer relate
55	15	75.0	13500	24	ABK10902	Human breast cancer
56	15	75.0	13500	24	ABN5767	Gene #2265 used to
57	15	75.0	13500	24	ABL62875	Breast cancer rela
58	15	75.0	13500	24	ABK63093	Breast cancer rela
59	15	75.0	13500	24	ABK64391	Stomach cancer rel
60	15	75.0	13500	24	ABK64745	Lung cancer relate
61	15	75.0	13500	24	ABL65822	Lung cancer relate
62	15	75.0	13500	24	ABL66491	Lung cancer relate
63	15	75.0	13500	24	ABL68813	Kidney cancer rela
64	15	75.0	13646	22	AAK97857	Human neuroblastom
65	15	75.0	30803	22	AAK68410	Human immune/haema
66	15	75.0	59747	22	ABO88209	Human osteoblast d
67	15	75.0	121162	21	AAK66568	Human kinesin-like
68	15	75.0	786431	24	ABO74964	Human kinase prote
69	15	75.0	4403765	22	AAI99683	Mycobacterium tube
70	15	75.0	4411529	22	AAI99682	Mycobacterium tube
71	14	70.0	20	24	ABK03748	Human RECOQ5 inhib
72	14	70.0	20	24	AAK97578	Murine SAG1 gene-s
73	14	70.0	20	24	AAK97581	Murine SAG1 gene-s
74	14	70.0	351	20	AAK40835	Human secreted pro
75	14	70.0	351	21	AAK00515	Human secreted pro
76	14	70.0	363	22	AAK59288	Human immune/haema
77	14	70.0	378	22	AAK74931	Human immune/haema
78	14	70.0	378	22	AAK74932	Human immune/haema
79	14	70.0	403	20	AAV88657	EST clone HE111.
80	14	70.0	451	22	AAK60454	Human immune/haema
81	14	70.0	463	22	AAK92885	Human CDNA 3'-end

C 82	14	70.0	692	22	AAK91798	Human CDNA 5'-end
C 83	14	70.0	692	22	AAK93918	Human CDNA clone i
C 84	14	70.0	817	20	AAZ40839	Secreted protein E
C 85	14	70.0	869	22	AAH98832	Human EST-derived
C 86	14	70.0	1065	22	AA840959	CDNA encoding novel
C 87	14	70.0	1108	20	AAZ31913	Human helicase, Re
C 88	14	70.0	1150	24	AAH68765	Human polynucleoti
C 89	14	70.0	1447	21	AAD00615	Human membrane tra
C 90	14	70.0	1467	22	AA159697	Human polynucleoti
C 91	14	70.0	1566	22	AA157911	Human polynucleoti
C 92	14	70.0	1637	23	ABL05105	Drosophila melanop
C 93	14	70.0	1740	22	AAK94445	Human full-length
C 94	14	70.0	1749	22	AAF87638	Human RecQ3 gamma
C 95	14	70.0	1799	25	ABZ55283	Human GPC polynuc
C 96	14	70.0	1832	11	AAQ01762	CDNA sequence of i
C 97	14	70.0	2170	24	AB199214	Mouse ischaemic c
C 98	14	70.0	2231	22	AAK94286	Human full-length
C 99	14	70.0	2579	20	AAZ33983	Human PRO705 nucle
C 100	14	70.0	2579	21	AA878482	Human PRO705 (UNO
C 101	14	70.0	2579	22	AA845972	Human DNA encoding
C 102	14	70.0	2579	25	ACR57730	Human PRO705 CDNA
C 103	14	70.0	2579	25	ABX96820	Human CDNA encodi
C 104	14	70.0	2579	25	ABX96702	Novel human secret
C 105	14	70.0	2579	25	ACA05747	Human secreted/tir
C 106	14	70.0	2579	25	ABX92355	CDNA encoding huma
C 107	14	70.0	2579	25	ABX97791	Human PRO polynuci
C 108	14	70.0	2579	25	ABX78575	Human PRO polynuci
C 109	14	70.0	2579	25	ABX75588	Human CDNA encodi
C 110	14	70.0	2579	25	ABX76793	human PNO polynuc
C 111	14	70.0	2579	25	ABX16633	Human CDNA encodi
C 112	14	70.0	2638	25	ABZ25916	Human PWP structur
C 113	14	70.0	2644	25	ABZ52821	Human GPC polynuc
C 114	14	70.0	2661	11	AAQ03458	Sequence encoding
C 115	14	70.0	2731	20	AAZ00683	Human GRC6 DNA, H
C 116	14	70.0	2760	25	ABZ59282	Human GPC polynuci
C 117	14	70.0	3707	20	AAZ31910	Human helicase, Re
C 118	14	70.0	3715	22	AAF87637	Human RecQ3 alpha
C 119	14	70.0	3822	23	ABG05104	Drosophila melanog
C 120	14	70.0	6401	25	ACG50236	Breast cancer assoc

ALIGNMENTS

```

RESULT 1
ABA00272
ID ABA00272 standard; DNA; 20 BP.
XX
AC ABA00272;
XX
DT 29-NOV-2002 (first entry)
XX
DE EBNA 2 sense primer.
XX
KM Primer; amplify; PCR; probe; detection; Epstein-Barr virus; EBV; ss.
XX
OS Epstein-barr virus.
XX
NC_020026484.2.
XX
PD 22-AUG-2002.
XX
PF 13-FEB-2002; 2002WC-US04339.
XX
PR 13-FEB-2001; 2001US-268439P.
XX
PA (CHIL-) CHILDRENS HOSPITAL RES FOUND.
PI Witte DP, Groen PA;
DR WPI; 2002-667015/71.
XX
New compositions comprising nucleic acid sequences which specifically

```

XX	hybridizes to Epstein-Barr virus (EBV) nucleic acid, for detecting EBV
PT	in clinical specimens to determine patients at high risk of to
PT	developing EBV infections -
XX	
PS	Claim 1, Page 44; 59pp; English.
XX	
CC	The sequences given in ABA00266-75 are primers and probes which were
CC	used in the compositions of the invention for the detection of
CC	Epstein-Barr virus (EBV). The compositions comprise at least one
CC	purified and isolated oligonucleotide consisting of a nucleic acid
CC	sequence which complements and specifically hybridizes to EBV nucleic
CC	acid. The oligonucleotide sequences and compositions comprising them
CC	are useful for detecting EBV in clinical specimens to determine
CC	patients who are at high risk to develop serious and costly medical
CC	complications, and allow for better clinical management of these
CC	patients by earlier recognition of their infection status. The
CC	oligonucleotide sequences may also be used to amplify EBV DNA
CC	sequences. The use of the oligonucleotide sequences in the assay for
CC	detecting EBV has a broad dynamic range of detection from less than
CC	10-100000000 copies of EBV DNA, is less labour intensive requiring only
CC	one reaction tube for the EBV determination, highly sensitive, accurate
CC	and has a rapid turn around time with assays that are completed, and
CC	including amplification, probe specific hybridization, and calculation
CC	of copy number in less than 1 hour. The method may be adapted to
CC	automated systems.
XX	
SO	Sequence 20 BP; 7 A; 4 C; 7 G; 2 T; 0 other;
XX	
Query Match	100.0%; Score 20; DB 24; Length 20;
Best Local Similarity	100.0%; Freq. No. 0.041;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps	
QY	1 AGGGATGCTCGACACAAAGA 20
Db	1 AGGGATGCTCGACACAAAGA 20
RESULT 2	
AAH03059	ID AAH03059 standard; DNA; 20 BP.
XX	
AC	AAH03059;
XX	
DT	15-JUN-2001 (first entry)
XX	
DE	Microorganism detection method related oligonucleotide SEQ ID NO: 83.
XX	
KM	Microorganism identification; pathogen; DNA sequencing; HLA type;
XX	b1-directional sequencing; infection; mutation detection; PCR primer; ss
XX	Undentified.
XX	
OS	US6214555-B1.
XX	
PD	10-APR-2001.
XX	
PF	13-MAY-1999; 99US-0311260.
XX	
PR	01-MAY-1996; 96US-0640672.
XX	19-JUN-1996; 96US-0684498.
PR	22-FEB-1997; 97US-0807138.
XX	20-JAN-1998; 98US-0009483.
PR	
XX	(VIST-) VISIBLE GENETICS INC.
PA	
XX	Leushner J, Hui M, Dunn JW, Lacroix J;
PI	WPI; 2001-289718/30.
XX	
DR	
XX	Composition for detecting microorganisms, comprising deoxynucleotide
PT	triphosphates, dideoxynucleotide triphosphate, and thermostable
XX	polymerase to incorporate dideoxynucleotide triphosphate into extending
PT	polymer -

XX PS Disclosure; Column 63; 62pp; English.

CC The present invention provides a composition containing 4 dNTPs and at least one ddNTP and a thermally stable polymerase which incorporates ddNTPs into an extending nucleic acid polymer at a rate of not less than 0.4 times the rate of dNTP incorporation. This can be used with the PCR primers provided in the invention to detect the presence of microorganisms, such as Chlamydia trachomatis, HIV or human papillomavirus, in a sample. In addition, it can be used to detect mutations in a specific gene, to determine HLA type, and to produce sequencing fragments for further study.

XX Sequence 20 BP; 8 A; 3 C; 6 G; 2 T; 0 other;

QY Query Match 90.0%; Score 18; DB 22; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 3 GGATGCTGACACAGA 20
3 GGATGCTGACACAGA 20

QY
DB

RESULT 3
AAQ2134/c
ID AAQ2134 standard; DNA; 20 BP.

XX AAQ2134;

AC 25-MAR-2003 (updated)

DT 03-JUL-1992 (first entry)

XX Primer D corresp. to region conserved between EBV type A and B.

DE Epstein Barr Virus; polymerase chain reaction; ss.

XX Synthetic.

OS

XX JP04027399-A
XX 30-JAN-1992

XX 23-MAY-1990; 90JP-0133618.

XX 23-MAY-1990; 90JP-0133618.

XX (SHIO) SHIONOGI & CO LTD.

XX WPI; 1992-084781/11.

DR Simultaneous type detection of Epstein-Barr virus and DNA sequence -
PT in which specimen is amplified using primer and PCR process, and
PT then amplified DNA is detected

XX Disclosure; Page 4; 6pp; Japanese.

XX PCR primer D is based on nucleotides 1816-1835 of EBV
CC type A and is used with primer A, B, C or E (see AAQ2134-Q22000 and
CC AAQ2132-Q22133) to amplify a fragment of the EBV genome. The primer
CC sequences are common to viral types A and B but the sequence which
CC is amplified differs between types A and B. Thus, by determining the
CC mol. wt. of the amplified intervening region (e.g. by electrophoresis),
CC it is possible to distinguish which EBV type is present in the sample.
CC See AAQ21396-Q22000 and AAQ22132-Q22139
CC (updated on 23-MAR-2003 to correct PA field.)

XX Sequence 20 BP; 2 A; 6 C; 6 G; 6 T; 0 other;

QY Query Match 85.0%; Score 17; DB 13; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGATGCTGACACAGA 20
DB 20 GGATGCTGACACAGA 4

RESULT 4
AAK68955
ID AAK68955 standard; DNA; 652 BP.

XX AAK68955;

AC

XX 06-NOV-2001 (first entry)

DT

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23767.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.


```

FH Key Location/Qualifiers
FT 1..1452
FT CDS /*tag= a
FT /product= "alpha-amylase"
FT /EC number= "3.2.1.1"
FT /partial
FT /note= "CDS lacks an initiation codon"
XX
XX WO2003014358-A2.
XX
XX 20-FEB-2003.
XX
XX 27-JUL-2002; 2002WO-EP08391.
XX
XX 07-AUG-2001; 2001DE-1038753.
XX
XX (HENK ) HENKEL KGAA.
XX
XX Kottwitz B, Breves R, Maurer K;
XX
XX WPI: 2003-278480/27.
XX
XX P-PSDB; ABP60488.
XX
XX Washing and cleaning composition, useful for laundry and hard surface
XX cleaning, contains hybrid amylase derived from two Bacillus enzymes -
XX
XX Claim 46; Page 89-91; 118pp; German.
XX
XX The invention relates to a washing and cleaning composition (A)
XX containing an amylolytic hybrid protein (I) containing sequences from the
XX alpha-amylases (EC 3.2.1.1) of Bacillus amyloliquefaciens and B.
XX licheniformis. (A) are used for cleaning textiles (by hand or machine) or
XX hard surfaces (metal, glass, plastic etc.), especially in dishwashing
XX machines. (I) can also be used to release other components of the
XX compositions from starch-based capsules. The present sequence is that of
XX the Bacillus amyloliquefaciens alpha-amylase encoding DNA of the
XX invention.
XX
XX Sequence 1452 BP; 426 A; 291 C; 388 G; 347 T; 0 other;
XX
XX Query Match 80.0%; Score 16; DB 25; Length 1452;
XX Best Local Similarity 100.0%; Pred. No. 8;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 GGATGCTGGACACAA 18
XX 950 GGATGCTGGACACAA 935
XX
XX RESULT 6
XX ABZ59269/c
XX ID ABZ59269 standard; DNA; 1452 BP.
XX
XX AC ABZ59269;
XX
XX 10-MAY-2003 (first entry)
XX
XX Bacillus alpha-amylase fusion protein ALA34-84 encoding DNA SEQ ID NO 13.
XX
XX Bacillus licheniformis; Bacillus amyloliquefaciens; alpha-amylase;
XX washing; cleaning; textile; dishwashing machine; starch-based capsule;
XX gene; ds.
XX
XX Chimeric - Bacillus licheniformis.
XX OS Chimeric - Bacillus amyloliquefaciens.
XX
XX Key Location/Qualifiers
XX CDS 1..1452
XX /*tag= a
XX /product= "alpha-amylase"
XX /EC number= "3.2.1.1"
XX /partial
XX /note= "CDS lacks an initiation codon"
XX
```

```

XX
XX PN WO2003014358-A2.
XX
XX 20-FEB-2003.
XX
XX 27-JUL-2002; 2002WO-EP08391.
XX
XX 07-AUG-2001; 2001DE-1038753.
XX
XX (HENK ) HENKEL KGAA.
XX
XX Kottwitz B, Breves R, Maurer K;
XX
XX WPI: 2003-278480/27.
XX
XX P-PSDB; ABP60493.
XX
XX Washing and cleaning composition, useful for laundry and hard surface
XX cleaning, contains hybrid amylase derived from two Bacillus enzymes -
XX
XX Disclosure; Page 107-109; 118pp; German.
XX
XX The invention relates to a washing and cleaning composition (A)
XX containing an amylolytic hybrid protein (I) containing sequences from the
XX alpha-amylases (EC 3.2.1.1) of Bacillus amyloliquefaciens and B.
XX licheniformis. (A) are used for cleaning textiles (by hand or machine) or
XX hard surfaces (metal, glass, plastic etc.), especially in dishwashing
XX machines. (I) can also be used to release other components of the
XX compositions from starch-based capsules. The present sequence is that of
XX the Bacillus amyloliquefaciens alpha-amylase
XX fusion protein encoding DNA of the invention.
XX
XX Sequence 1452 BP; 416 A; 290 C; 398 G; 348 T; 0 other;
XX
XX Query Match 80.0%; Score 16; DB 25; Length 1452;
XX Best Local Similarity 100.0%; Pred. No. 8;
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 GGATGCTGGACACAA 18
XX 950 GGATGCTGGACACAA 935
XX
XX RESULT 7
XX ABZ59270/c
XX ID ABZ59270 standard; DNA; 1458 BP.
XX
XX AC ABZ59270;
XX
XX 10-MAY-2003 (first entry)
XX
XX Bacillus alpha-amylase fusion protein LAL19-433 encoding DNA SEQ ID 15.
XX
XX Bacillus licheniformis; Bacillus amyloliquefaciens; alpha-amylase;
XX washing; cleaning; textile; dishwashing machine; starch-based capsule;
XX gene; ds.
XX
XX Chimeric - Bacillus licheniformis.
XX OS Chimeric - Bacillus amyloliquefaciens.
XX
XX Key Location/Qualifiers
XX CDS 1..1458
XX /*tag= a
XX /product= "alpha-amylase"
XX /EC number= "3.2.1.1"
XX /partial
XX /note= "CDS lacks an initiation codon"
XX
XX PN WO2003014358-A2.
XX
XX 20-FEB-2003.
XX
XX 27-JUL-2002; 2002WO-EP08391.
XX
```

```

PR 07-AUG-2001; 2001DE-1038753.
PA (HENK ) HENKEL KGAA.
PI Kottwitz B, Breves R, Maurer K;
DR WPI; 2003-278480/27.
DR P-PSDB; AAB60494.
PT Washing and cleaning composition, useful for laundry and hard surface
PT cleaning, contains hybrid amylase derived from two Bacillus enzymes
PS Disclosure; Page 11-113; 118pp; German.
XX
XX The invention relates to a washing and cleaning composition (A)
XX containing an amylolytic hybrid protein (I) containing sequences from the
XX alpha-amylases (EC 3.2.1.1) of Bacillus amyloliquefaciens and B.
XX licheniformis. (A) are used for cleaning textiles (by hand or machine) or
XX hard surfaces (metal, glass, plastic etc.), especially in dishwashing
XX machines. (I) can also be used to release other components of the
XX compositions from starch-based capsules. The present sequence is that of
XX Bacillus licheniformis and Bacillus amyloliquefaciens alpha-amylase
XX fusion protein encoding DNA of the invention.
SQ Sequence 1458 BP; 422 A; 292 C; 395 G; 349 T; 0 other;
Query Match 80.0%; Score 16; DB 25; Length 1458;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGGACACAA 18
DB 956 GGATGCTGGACACAA 941
RESULT 8
AAA60576/c
ID AAA60576 standard; DNA; 1545 BP.
XX
XX AAA60576;
AC
XX 19-OCT-2000 (first entry)
DT
XX
XX Bacillus amyloliquefaciens clone number 21 SEQ ID NO:2.
DE
XX
XX Bacillus amyloliquefaciens; alpha-amylase; thermostable; bread; ds.
KM
XX
XX Bacillus amyloliquefaciens.
OS
XX
XX Key Location/Qualifiers
FH CDS 1..1545
FT /*tag= a
FT sig_peptide 1..93
FT /*tag= b
FT mat_peptide 94..1542
FT /*tag= c
XX
XX JP2000135093-A.
PN
XX
XX 16-MAY-2000.
PD
XX
XX 20-AUG-1999; 99JP-0234813.
PF
XX
XX 24-AUG-1998; 98JP-0237839.
PR
XX
XX (DAIW ) DAIWA KASEI KK.
PA
XX
XX WPI; 2000-403584/35.
DR
XX
XX P-PSDB; AAB12429.
PT Novel thermostable alpha-amylase, useful for improving the preparation
PT of bread, comprises alpha-amylase activity with less than 80% activity
PT after treatment at 65degreesC for 30 min -
XX
XX

```

```

XX
XX Claim 10; Page 11; 22pp; Japanese.
PS
XX
XX The present invention describes a thermostable alpha-amylase (I)
XX comprising the sequence given in AAB12433 (A) or deletions, replacements
XX or insertions of one or more amino acid(s) in the sequence and
XX alpha-amylase activity with less than 80% activity after treatment at
XX 65 plus degrees Celsius for 30 minutes. Also described are: (1) DNAs
XX encoding (I); (2) DNAs containing one of 4 nucleotide sequences
XX comprising 1545 base pairs (see AAA60576 to AAA60579); (3) expression
XX vectors containing the above mentioned DNAs; (4) host cells containing
XX the above mentioned expression vectors; (5) preparation of a polypeptide
XX having alpha-amylase activity by culture of the recombinant host cells;
XX and (6) preparation of bread including a process to add (I) to the
XX dough. (I) is useful in the preparation of bread. (I) improves the soft
XX volume of the bread and prevents it aging. The present sequence encodes
XX a Bacillus amyloliquefaciens clone number 21 protein, which is used in
XX the exemplification of the present invention.
SQ Sequence 1545 BP; 454 A; 311 C; 406 G; 374 T; 0 other;
Query Match 80.0%; Score 16; DB 21; Length 1545;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGGACACAA 18
DB 1043 GGATGCTGGACACAA 1028
RESULT 9
AAA60577/c
ID AAA60577 standard; DNA; 1545 BP.
XX
XX AAA60577;
AC
XX 19-OCT-2000 (first entry)
DT
XX
XX Bacillus amyloliquefaciens clone number 22 SEQ ID NO:4.
DE
XX
XX Bacillus amyloliquefaciens; alpha-amylase; thermostable; bread; ds.
KM
XX
XX Bacillus amyloliquefaciens.
OS
XX
XX Key Location/Qualifiers
FH CDS 1..1545
FT /*tag= a
FT sig_peptide 1..93
FT /*tag= b
FT mat_peptide 94..1542
FT /*tag= c
XX
XX JP2000135093-A.
PN
XX
XX 16-MAY-2000.
PD
XX
XX 20-AUG-1999; 99JP-0234813.
PF
XX
XX 24-AUG-1998; 98JP-0237839.
PR
XX
XX (DAIW ) DAIWA KASEI KK.
PA
XX
XX WPI; 2000-403584/35.
DR
XX
XX P-PSDB; AAB12430.
PT Novel thermostable alpha-amylase, useful for improving the preparation
PT of bread, comprises alpha-amylase activity with less than 80% activity
PT after treatment at 65degreesC for 30 min -
XX
XX
XX Claim 10; Page 13; 22pp; Japanese.
PS
XX
XX The present invention describes a thermostable alpha-amylase (I)
XX comprising the sequence given in AAB12433 (A) or deletions, replacements
XX

```

or insertions of one or more amino acid(s) in the sequence and
CC alpha-amylase activity with less than 80% activity after treatment at
CC 65 plus degrees Celsius for 30 minutes. Also described are: (1) DNAs
CC encoding (1); (2) DNAs containing one of 4 nucleotide sequences
CC comprising 1545 base pairs (see AAA60576 to AAA60579); (3) expression
CC vectors containing the above mentioned DNAs; (4) host cells containing
CC the above mentioned expression vectors; (5) preparation of a polypeptide
CC having alpha-amylase activity by culture of the recombinant host cells;
CC and (6) preparation of bread including a process to add (1) to the
CC dough. (1) is useful in the preparation of bread. (1) improves the soft
CC volume of the bread and prevents it aging. The present sequence encodes
CC a Bacillus amyloliquefaciens clone number 22 protein, which is used in
CC the exemplification of the present invention.
XX
SQ Sequence 1545 BP; 454 A; 309 C; 406 G; 376 T; 0 other;
Query Match 80.0%; Score 16; DB 21; Length 1545;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 GGATGCTGACACAA 18
DB 1043 GGATGCTGACACAA 1028
RESULT 10
AAA60578/c
ID AAA60578 standard; DNA; 1545 BP.
XX
AC AAA60578;
XX
DT 19-OCT-2000 (first entry)
XX
DE Bacillus amyloliquefaciens clone number 24 SEQ ID NO:6.
XX
KW Bacillus amyloliquefaciens; alpha-amylase; thermostable; bread; ds.
XX
OS Bacillus amyloliquefaciens.
XX
XX Key Location/Qualifiers
FH CDS 1..1545
FT sig_peptide 1..93
FT mat_peptide 94..1542
FT /*tag= b
FT /*tag= c
XX
PN JP2000135093-A.
XX
XX 16-MAY-2000.
XX
PD 20-AUG-1999; 99JP-0234813.
XX
PF 24-AUG-1998; 98JP-0237839.
XX
PR (DAIW) DAIWA KASEI KK.
XX
PA WPI; 2000-403584/35.
XX
DR P-PSDB; AAB12431.
XX
DR Novel thermostable alpha-amylase, useful for improving the preparation
PT of bread, comprises alpha-amylase activity with less than 80% activity
PT after treatment at 65degreesc for 30 min -
XX
XX Claim 10; Page 15-16; 22pp; Japanese.
XX
XX The present invention describes a thermostable alpha-amylase (1)
CC comprising the sequence given in AAB12433 (A) or deletions, replacements
CC or insertions of one or more amino acid(s) in the sequence and
CC alpha-amylase activity with less than 80% activity after treatment at
CC 65 plus degrees Celsius for 30 minutes. Also described are: (1) DNAs
CC encoding (1); (2) DNAs containing one of 4 nucleotide sequences
CC comprising 1545 base pairs (see AAA60576 to AAA60579); (3) expression
CC vectors containing the above mentioned DNAs; (4) host cells containing
CC the above mentioned expression vectors; (5) preparation of a polypeptide
CC having alpha-amylase activity by culture of the recombinant host cells;
CC and (6) preparation of bread including a process to add (1) to the
CC dough. (1) is useful in the preparation of bread. (1) improves the soft

vectors containing the above mentioned DNAs; (4) host cells containing
CC the above mentioned expression vectors; (5) preparation of a polypeptide
CC having alpha-amylase activity by culture of the recombinant host cells;
CC and (6) preparation of bread including a process to add (1) to the
CC dough. (1) is useful in the preparation of bread. (1) improves the soft
CC volume of the bread and prevents it aging. The present sequence encodes
CC a Bacillus amyloliquefaciens clone number 24 protein, which is used in
CC the exemplification of the present invention.
XX
SQ Sequence 1545 BP; 455 A; 309 C; 405 G; 376 T; 0 other;
Query Match 80.0%; Score 16; DB 21; Length 1545;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 GGATGCTGACACAA 18
DB 1043 GGATGCTGACACAA 1028
RESULT 11
AAA60579/c
ID AAA60579 standard; DNA; 1545 BP.
XX
AC AAA60579;
XX
DT 19-OCT-2000 (first entry)
XX
DE Bacillus amyloliquefaciens clone number 25 SEQ ID NO:8.
XX
KW Bacillus amyloliquefaciens; alpha-amylase; thermostable; bread; ds.
XX
OS Bacillus amyloliquefaciens.
XX
XX Key Location/Qualifiers
FH CDS 1..1545
FT sig_peptide 1..93
FT mat_peptide 94..1542
FT /*tag= b
FT /*tag= c
XX
PN JP2000135093-A.
XX
XX 16-MAY-2000.
XX
PD 20-AUG-1999; 99JP-0234813.
XX
PF 24-AUG-1998; 98JP-0237839.
XX
PR (DAIW) DAIWA KASEI KK.
XX
PA WPI; 2000-403584/35.
XX
DR P-PSDB; AAB12432.
XX
DR Novel thermostable alpha-amylase, useful for improving the preparation
PT of bread, comprises alpha-amylase activity with less than 80% activity
PT after treatment at 65degreesc for 30 min -
XX
XX Claim 10; Page 17-18; 22pp; Japanese.
XX
XX The present invention describes a thermostable alpha-amylase (1)
CC comprising the sequence given in AAB12433 (A) or deletions, replacements
CC or insertions of one or more amino acid(s) in the sequence and
CC alpha-amylase activity with less than 80% activity after treatment at
CC 65 plus degrees Celsius for 30 minutes. Also described are: (1) DNAs
CC encoding (1); (2) DNAs containing one of 4 nucleotide sequences
CC comprising 1545 base pairs (see AAA60576 to AAA60579); (3) expression
CC vectors containing the above mentioned DNAs; (4) host cells containing
CC the above mentioned expression vectors; (5) preparation of a polypeptide
CC having alpha-amylase activity by culture of the recombinant host cells;
CC and (6) preparation of bread including a process to add (1) to the
CC dough. (1) is useful in the preparation of bread. (1) improves the soft

CC volume of the bread and prevents it aging. The present sequence encodes
CC a *Bacillus amyloliquefaciens* clone number 25 protein, which is used in
CC the exemplification of the present invention.

XX Sequence 1545 BP; 454 A; 309 C; 406 G; 376 T; 0 other;

Query Match 80.0%; Score 16; DB 21; Length 1545;

Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTCGACACAA 18

DB 1043 GGATGCTCGACACAA 1028

RESULT 12

ID AAQ10322/C

XX AAQ10322;

AC AAQ10322;

DT 25-MAR-2003 (updated)

DT 05-APR-1991 (first entry)

DE EcoR1-BamH1 fragment of *pMatBac* encoding alpha-amylase gene.

XX pUCam4; ds.

XX *Bacillus amyloliquefaciens*.

XX Key Location/Qualifiers

FT CDS 138..1682

FT /*tag= a

XX EP409299-A.

XX 23-JAN-1991.

XX 27-JUN-1990; 90EP-0201704.

XX 29-JUN-1989; 89EP-0201732.

XX 27-JUN-1990; 90EP-0201704.

XX (KONN) GIST-BROCADES NV.

XX Vanejik JH, Quax WJ, Sanders JP;

XX WPI, 1991-023909/04.

XX P-PSDB; AAR10378; AAR10379.

XX Mutant enzyme having reduced stability - comprising modified

XX alpha-amylase which can be used in baking to improve loaf vol.

XX and crumb softness without over-dextrinisation

XX Claim 18; Fig 3; 26pp; English.

XX The wild type amylase sequence differs at Arg123-Cys, and optionally

XX at positions 113-4, 116, 123, 163-4, 166, 238, 316, 322, 345, 349,

XX 356, 366, 394 or 398. The mutant for is less stable under industrial

XX conditions, and may be used as active ingredient in bread making,

XX improving crumb softness and loaf volume without over dextrinisation

XX of starch.

XX (Updated on 25-MAR-2003 to correct PI field.)

CC Sequence 1972 BP; 565 A; 358 C; 518 G; 491 T; 0 other;

Query Match 80.0%; Score 16; DB 12; Length 1972;

Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTCGACACAA 18

DB 1180 GGATGCTCGACACAA 1165

RESULT 13
ID AAV02472/C

XX AAV02472;

AC AAV02472;

DT 11-MAY-1998 (first entry)

DE *Bacillus amyloliquefaciens* Termamyl-like alpha-amylase DNA.

XX Termamyl; alpha-amylase; enzyme engineering; protein engineering;

XX starch; liquefaction; saccharification; sweetener;

XX textile desizing; detergent additive; ss.

XX *Bacillus amyloliquefaciens*.

XX Key Location/Qualifiers

FT CDS 249..1793

FT /*tag= a

FT sig_peptide 249..341

FT /*tag= b

FT mat_peptide 342..1790

FT /*tag= c

FT /product= Termamyl-like alpha-amylase

XX WO9741213-A1.

XX 06-NOV-1997.

XX 30-APR-1997; 97WO-DK00197.

XX 08-NOV-1996; 96DK-0001263.

XX 30-APR-1996; 96DK-0000515.

XX 28-JUN-1996; 96DK-0000712.

XX 11-JUL-1996; 96DK-0000775.

XX (NOVO) NOVO-NORDISK AS.

XX Bisgard-Frantzen H, Borchert TV, Svendsen A;

XX WPI, 1997-549718/50.

XX P-PSDB; AAW31405.

XX Termamyl-like alpha-amylase variants with improved properties - e.g.

XX increased stability at low pH and low calcium, useful as detergent

XX additives and in industrial starch processing e.g. liquefaction

XX disclosure; Page 82-83; 101pp; English.

XX This DNA sequence includes a coding region for a Termamyl-like

XX alpha-amylase (see AAW31405) of *Bacillus amyloliquefaciens*. The

XX invention relates to novel variants (mutants) of Termamyl-like

XX alpha-amylases that have alpha-amylase activity and exhibit an

XX alteration in at least one property selected from: substrate

XX specificity; binding or cleavage pattern; thermal stability;

XX pH/activity or pH/stability profile; stability towards oxidation;

XX Ca2+ dependency and specific activity. The variant has one or

XX more mutations from those listed in the specification in relation

XX to *Bacillus licheniformis* Termamyl (see AAW31404). Also claimed are

XX constructs comprising DNA encoding the variant, and recombinant

XX expression vectors and transformed cells containing the DNA.

XX The Termamyl-like alpha-amylase variant is useful as a detergent

XX additive and can also be used in industrial starch processing e.g.

XX liquefaction (claimed) or saccharification to produce sweeteners.

XX and in textile desizing (claimed).

XX Sequence 2083 BP; 610 A; 401 C; 543 G; 529 T; 0 other;

Query Match 80.0%; Score 16; DB 18; Length 2083;

Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
 DB 1291 GGATGCTGGACACAA 1276

RESULT 14
 ID AAQ8067/c
 AAQ8067 standard; DNA; 2084 BP.

XX AAQ8067;
 XX
 XX 25-MAR-2003 (updated)
 DT 01-DEC-1995 (first entry)

XX Bacillus amyloliquefaciens alpha amylase coding sequence.

XX Alpha amylase; variant; enzyme; detergent; additive; dishwashing;
 KW washing; Bacillus licheniformis; Bacillus amyloliquefaciens;
 KW Bacillus stearothermophilus; dyeing; bleaching; scouring; textile;
 KW thermostable; ss.

XX Bacillus amyloliquefaciens.

XX Key Location/Qualifiers
 FT CDS 250..1794
 FT /*tag= a
 FT /product= Alpha amylase.

FT sig_peptide 250..342
 FT /*tag= b
 FT mat_peptide 343..1791
 FT /*tag= c

XX MO9510603-A1.

XX 20-APR-1995.

XX 05-OCT-1994; 94WO-DK00370.

XX 08-OCT-1993; 93DK-0001133.
 PR 02-FEB-1994; 94DK-0000140.

XX (NOVO) NOVO-NORDISK AS.

XX Borchert TV, Bisgard-frantzen H, Svendsen A, Thellersen M;
 PI Van Der Zee P;

XX WPI; 1995-161790/21.
 DR P-PSDB; AAR72448.

XX New Bacillus derived alpha-amylase variants - having amino acid
 PT modifications to improve washing and/or dishwashing performance
 XX

PS Disclosure; Page 72-73; 105pp; English.

XX Variant alpha amylase enzymes which have improved washing and/or
 CC as detergent additives. The enzymes have one or more amino acid
 CC residues added, deleted or substituted. The variants can also be
 CC used for textile desizing prior to scouring, bleaching and dyeing.
 CC The variants have improved thermostability, acid/alkaline stability;
 CC low temperature optimum; pH optimum; higher hydrolysis velocity and
 CC improved tolerance to other composition constituents, e.g. oxidation
 CC agents.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 2084 BP; 610 A; 401 C; 544 G; 529 T; 0 other;

Query Match 80.0%; Score 16; DB 16; Length 2084;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18

DB 1292 GGATGCTGGACACAA 1277

RESULT 15
 ID AAQ95032/c
 AAQ95032 standard; DNA; 2084 BP.

XX AAQ95032;
 XX
 XX 17-JAN-1996 (first entry)

XX Bacillus amyloliquefaciens alpha amylase gene.

XX Alpha amylase; oxidation; desizing; bleaching; scouring; fabric;
 KW starch; thermostable; methionine; Bacillus licheniformis;
 KW Bacillus amyloliquefaciens; Bacillus stearothermophilus; ss.

XX Bacillus amyloliquefaciens.

XX Key Location/Qualifiers
 FT 5'UTR 1..249
 FT /*tag= a
 FT CDS 250..1794
 FT /*tag= b
 FT /product= Alpha amylase.

FT sig_peptide 250..342
 FT /*tag= c
 FT mat_peptide 343..1791
 FT /*tag= d
 FT 3'UTR 1795..2084
 FT /*tag= e

XX MO9521247-A1.

XX 10-AUG-1995.

XX 05-OCT-1994; 94WO-DK00371.

XX 02-FEB-1994; 94DK-0000141.

XX (NOVO) NOVO-NORDISK AS.

XX Marcher D, Nilsson TE, Pedersen HH, Toft AH;
 PI WPI; 1995-283767/37.

XX DR P-PSDB; AAR78268.

XX Use of an oxidation stable alpha-amylase - for simultaneous desizing
 FT and bleaching or scouring of fabrics contg. starch or starch derivs.
 XX

PS Disclosure; Page 22-24; 37pp; English.

XX Oxidation stable alpha amylases can be used for the simultaneous
 CC desizing and bleaching or scouring of a fabric comprising starch or
 CC starch derivatives. They exhibit a better heat stability,
 CC especially in the presence of oxidizing agents. They are obtained
 CC from a parent alpha amylase by replacing one or more methionine
 CC residues with any amino acid different from Cys or Met, preferably
 CC Leu, Thr, Ala, Gly, Ser, Ile or Asp. The parent alpha amylase is
 CC derived from a Bacillus species. This sequence encodes the wild
 CC type (unmodified) alpha amylase.

XX Sequence 2084 BP; 610 A; 401 C; 544 G; 529 T; 0 other;

Query Match 80.0%; Score 16; DB 16; Length 2084;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
 DB 1292 GGATGCTGGACACAA 1277

RESULT 16
 AAX59681/c
 ID AAX59681 standard; DNA; 2084 BP.
 AC AAX59681;
 DT 22-JUL-1999 (first entry)
 DE DNA encoding a termamyl-like alpha-amylase protein.
 KM Termamyl-like; alpha-amylase; variant; washing; dishwashing;
 KM production; sweetener; ethanol; starch; textile desizing;
 KM starch liquefaction; saccharification process; ss.
 OS Bacillus sp.
 PN WO9923211-A1.
 PD 14-MAY-1999.
 PF 30-OCT-1998; 98WO-DK00471.
 PR 14-JUL-1998; 98DK-0000936.
 PR 30-OCT-1997; 97DK-0001240.
 PA (NOVO) NOVO-NORDISK AS.
 PI Andersen C, Borchert TV, Kjaerulff S, Nielsen BR;
 PI Nissen TL, Svendsen A;
 DR WPI; 1999-326987/27.
 PT New Termamyl-like alpha-amylase variants
 PS Disclosure; Page 96-97; 115pp; English.
 CC The specification describes termamyl-like alpha-amylase variants that
 CC have altered amino acid sequences to improve properties. The variants
 CC are produced by creating one or more of the following mutations in
 CC amino acid sequence of the parent termamyl-like alpha-amylase: T141,
 CC K142, R143, D144, R145, P146, G147, R148, G149, Q174, R181, G182, D183,
 CC G184, K185, A186, W189, S193, N195, H107, K108, G109, D166, D168,
 CC Q169, S170, R171, Q172, F173, F267, W268, K269, N270, D271, L272, G273,
 CC A274, L275, K311, E346, K385, G456, N457, K458, P459, G460, T461, V462,
 CC T463. The variants can be used for washing and/or dishwashing. They can
 CC also be used in the production of sweeteners and ethanol from starch,
 CC and/or for textile desizing, and in starch liquefaction and/or
 CC saccharification processes. The present sequence encodes an amylase
 CC that can function as the parent sequence in the production of the
 CC variants of the invention.
 SQ Sequence 2084 BP; 610 A; 401 C; 544 G; 529 T; 0 other;
 QY 3 GGATGCTGGACACAA 18
 DB 1292 GGATGCTGGACACAA 1277
 DE
 AC AAX57596/c
 ID AAX57596 standard; DNA; 2084 BP.
 AC AAX57596;
 DT 16-JUL-1999 (first entry)
 DE Wild type Termamyl (RM)-like alpha-amylase coding sequence #5.

KM Variant; Termamyl; alpha-amylase; mutation; Bacillus; detergent;
 KM dishwashing; laundry; textile; desizing; starch liquefaction;
 KM sweetener; ethanol; ss.
 OS Bacillus sp.
 PN WO9919467-A1.
 PD 22-APR-1999.
 PF 13-OCT-1998; 98WO-DK00444.
 PR 13-OCT-1997; 97DK-0001172.
 PA (NOVO) NOVO-NORDISK AS.
 PI Bisgard-Frantzen H, Borchert TV, Svendsen A;
 DR WPI; 1999-277632/23.
 PT Variant alpha-amylases - useful as detergents or for textile
 PT desizing or starch liquefaction
 PS Disclosure; Page 81-82; 93pp; English.
 CC This sequence represents the coding sequence for a parent sequence
 CC used to generate new variants of a Termamyl-like alpha-amylase with
 CC alpha-amylase activity. The variants comprise mutations in 2-6
 CC regions/positions relative to an alpha-amylase from either of two
 CC Bacillus species in WO9926397, B. stercorophilus, B. licheniformis,
 CC B. amyloliquefaciens or Bacillus sp. #707. The alpha-amylase variants
 CC are detergent additives for use in detergents for dishwashing, manual
 CC or automatic laundry. The variants can also be used for textile desizing
 CC or starch liquefaction (e.g. for production of sweeteners or ethanol).
 SQ Sequence 2084 BP; 610 A; 401 C; 544 G; 529 T; 0 other;
 QY 3 GGATGCTGGACACAA 18
 DB 1292 GGATGCTGGACACAA 1277
 DE
 AC AAA48484/c
 ID AAA48484 standard; DNA; 2084 BP.
 AC AAA48484;
 DT 04-SEP-2000 (first entry)
 DE Bacillus amyloliquefaciens Termamyl-like alpha-amylase DNA sequence.
 KM Bacillus; alpha-amylase; washing; textile desizing;
 KM starch liquefaction; saccharification; muten; mutant;
 KM enzyme stability; hybrid; ss.
 OS Bacillus amyloliquefaciens.
 PN WO200029560-A1.
 PD 25-MAY-2000.
 PF 16-NOV-1999; 99WO-DK00628.
 FT Key Location/Qualifiers
 FT CDS 343..1794
 FT /*tag= a
 FT /product= "Termamyl-like alpha-amylase"
 FT /partial

```

XX PR 16-NOV-1998; 98DK-0001495.
XX PA (NOVO ) NOVO-NORDISK AS.
XX PI Svendsen A, Kjaerulff S, Bisgard-Frantzen H, Andersen C;
XX DR WPI: 2000-38777/33.
XX DR P-PSDB; AAY9606.
XX PS
XX PS Disclosure; Page 75-76; 80pp; English.
XX CC The present sequence encodes a parent alpha-amylase from which mutants
XX CC with increased stability at acidic pH, low calcium concentration and high
XX CC temperatures have been derived. The sequence was isolated from a Bacillus
XX CC amyloliquefaciens genomic DNA library. A variant may contain mutations in
XX CC one or more solvent exposed amino acid residues to increase the overall
XX CC hydrophobicity of the enzyme or the overall number of methyl groups in
XX CC the side chains of exposed residues may be increased. The mutations can
XX CC be incorporated by site-directed mutagenesis or by random mutagenesis. As
XX CC a result of their increased stability, the variants are suitable for the
XX CC industrial processing of starch, i.e. starch liquefaction and
XX CC saccharification. They may also be useful for washing, dishwashing and
XX CC textile desizing. Hybrid alpha-amylases comprising partial amino acid
XX CC sequences derived from two or more alpha-amylases have also been created
XX CC in order to increase enzyme stability.
SQ Sequence 2084 BP; 610 A; 401 C; 544 G; 529 T; 0 other;

Query Match 80.0%; Score 16; DB 21; Length 2084;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTTGACACAA 18
Db 1292 GGATGCTTGACACAA 1277

RESULT 19
ABL96211/c
ID ABL96211 standard; DNA; 2084 BP.
XX AC ABL96211;
XX AC
XX AC 19-AUG-2002 (first entry)
XX DT
XX DE Termamyl-like-alpha-amylases encoding sequence #5.
XX DE
XX DE Termamyl; alpha amylase; starch liquefaction; ethanol production;
XX DE textile desizing; detergent; enzyme; gene; ds.
XX DE
XX OS Bacillus amyloliquefaciens.
XX OS
XX FH Key Location/Qualifiers
XX FT CDS 343..1794
XX FT /tag= a
XX FT /product= "Termamyl-like alpha amylase"
XX FT /partial
XX FT /note= "no start codon"
XX PN
XX PN WO200210355-A2.
XX PD
XX PD 07-FEB-2002.
XX PF
XX PF 12-JUL-2001; 2001WO-DK00488.
XX PR
XX PR 01-AUG-2000; 2000DK-0001160.
XX PR 12-SEP-2000; 2000DK-0001354.
XX PR 10-NOV-2000; 2000DK-0001687.

```

```

PR 26-APR-2001; 2001DK-0000655.
XX CC
XX CC (NOVO ) NOVOZYMES AS.
XX CC
XX CC Thisted T, Kjaerulff S, Andersen C, Fuglsang CC;
XX CC WPI: 2002-28063/32.
XX CC DR P-PSDB; ABB76590.
XX CC
XX CC Variant of parent Termamyl-like alpha amylase, useful in detergent
XX CC compositions, for starch liquefaction, ethanol production, washing
XX CC and/or dish washing, and textile desizing
XX CC
XX PS Disclosure; Page 69-71; 90pp; English.
XX CC This invention relates to variants of a parent Termamyl-like
XX CC alpha-amylases. These are used for starch liquefaction, ethanol
XX CC production, detergent, and textile desizing. The amylases have altered
XX CC stability, particularly at high temperatures from 70-120plusoc and
XX CC low pH in the range from pH 4.0-6.0. The present sequence is a
XX CC termamyl-like-alpha-amylase encoding sequence.
SQ Sequence 2084 BP; 610 A; 401 C; 544 G; 529 T; 0 other;

Query Match 80.0%; Score 16; DB 24; Length 2084;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTTGACACAA 18
Db 1292 GGATGCTTGACACAA 1277

RESULT 20
ABL50568/c
ID ABL50568 standard; DNA; 2084 BP.
XX AC ABL50568;
XX AC
XX AC 19-JUN-2002 (first entry)
XX DT
XX DE B. amyloliquefaciens termamyl-like alpha-amylase encoding DNA SEQ ID:9.
XX DE
XX DE Bacillus; termamyl-like alpha-amylase; alpha-amylase; EC 3.2.1.1;
XX DE variant; mutant; enzyme; protein co-ordinate data; cleaning; detergent;
XX DE washing; sweetener; ethanol; starch; gene; ds.
XX DE
XX OS Bacillus amyloliquefaciens.
XX OS
XX FH Key Location/Qualifiers
XX FT CDS 343..1794
XX FT /tag= a
XX FT /EC_number= "3.2.1.1"
XX FT /product= "Termamyl-like alpha-amylase"
XX PN
XX PN WO200166712-A2.
XX PD
XX PD 13-SEP-2001.
XX PF
XX PF 07-MAR-2001; 2001WO-DK00144.
XX PR
XX PR 08-MAR-2000; 2000DK-0000376.
XX PR 15-MAR-2000; 2000US-189857P.
XX PR 23-FEB-2001; 2001DK-0000303.
XX PR 26-FEB-2001; 2001US-271382P.
XX PA
XX PA (NOVO ) NOVOZYMES AS.
XX PI
XX PI Andersen C, Borchert TV, Nielsen BR;
XX DR WPI: 2002-239612/29.
XX DR P-PSDB; ABB06937.

```

PT Novel variant of parent termamyl-like alpha-amylase useful as a
 PT component in washing and dishwashing compositions, for textile
 PT desizing, for starch liquefaction, and for producing sweeteners and
 PT ethanol from starch -
 PS Disclosure, Page 143-145; 153pp; English.
 XX
 XX The present invention describes a variant of a parent termamyl-like
 CC alpha-amylase (EC 3.2.1.1) (I) comprising an alteration at one or more
 CC positions of a group of 31 possible amino acid positions. The alteration
 CC in (I) may be at Arg28, Arg118, Asn174, Arg181, Gly182, Asp183, Gly184,
 CC Gly186, Trp189, Asn195, Met202, Tyr228, Asn239, Lys302, Ser303, Asn306,
 CC Arg310, Asn314, Arg320, His324, Glu345, Tyr396, Arg400, Trp439, Arg444,
 CC Asn445, Lys446, Glu449, Arg458, Asn471, or Asn484. (I) can be used for
 CC washing and/or dishwashing, textile desizing, and starch liquefaction.
 CC (I) is useful as a component in hard surface cleaning detergent
 CC composition, and for producing sweeteners and ethanol from starch.
 CC (I) has altered solubility, preferably increased solubility, in
 CC particular under washing, dish washing or hard surface cleaning
 CC conditions. The present sequence encodes a Bacillus amyloliquefaciens
 CC termamyl-like alpha-amylase which is used in the exemplification of
 CC the present invention.
 XX
 XX Sequence 2084 BP; 610 A; 401 C; 544 G; 529 T; 0 other;
 SO
 Query Match 80.0%; Score 16; DB 24; Length 2084;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 GGATGCTGGACACAA 18
 |||||
 DB 1292 GGATGCTGGACACAA 1277
 |||||
 RESULT 21
 AAI72215/c
 ID AAI72215 standard; cDNA; 2084 BP.
 XX
 AC AAI72215;
 XX
 DT 02-APR-2002 (first entry)
 XX
 DE Bacillus alpha amylase BAN cDNA.
 XX
 XX Alpha amylase; Bacillus; Termamyl-like; maltodextrin; glucose syrup;
 KM starch; food; feed; pharmaceutical; confectionery; candy;
 KM isotonic drink; bakery; cereal bar; ice cream; coffee whitener;
 KM salad dressing; cured meat; fermented meat; spice; ss.
 XX
 XX Bacillus amyloliquefaciens.
 OS
 XX Key Location/Qualifiers
 FH 343..1794
 FT CDS /*tag= a
 FT /*product= "Alpha-amylase"
 FT
 XX WO200196537-A2.
 PN
 XX 20-DEC-2001.
 PD
 XX 13-JUN-2001; 2001WO-DK00404.
 PF
 XX 14-JUN-2000; 2000DK-0000917.
 PR
 XX 20-JUN-2000; 2000US-212852P.
 XX
 XX (NOVO) NOVOZYMES AS.
 PA
 XX Nilsen BR, Weibye M;
 PI
 XX WPI; 2002-098064/13.
 DR
 XX P-PSDB; AAB47854.
 XX
 PT New modified alpha-amylase derived from the genus Bacillus and/or is a

PT Termamyl-like alpha-amylase, which has been pre-oxidized for producing
 PT maltodextrin or glucose syrup -
 PS Claim 7, Page 37-39; 47pp; English.
 XX
 XX The sequences given in AAI72211-16 encode modified alpha-amylases
 CC derived from the genus Bacillus. These alpha amylases are Termamyl-
 CC like alpha-amylase and they have been pre-oxidized. The alpha amylase
 CC is useful for producing a maltodextrin or glucose syrup, by treating
 CC starch with a pre-oxidized alpha-amylase until a product with a
 CC DE between 5-45 has been provided and/or a product with a
 CC molecular weight of between 5-30 kDa has been provided. The product
 CC comprises a maltodextrin with a DE of 18.5 and/or a maltodextrin with a
 CC molecular weight of 14-16 kDa. The alpha amylase is useful for producing
 CC a maltodextrin or glucose syrup, where the glucose syrup is useful as an
 CC ingredient in food, feed or pharmaceuticals. Glucose syrup is useful
 CC in confectionery such as candies, beverages such as isotonic drinks,
 CC bakery such as cereal bars, dairy and ice cream such as coffee
 CC whiteners, conventional foods such as salad dressings, and food
 CC ingredients and preparations such as cured meat, fermented meat, spices
 CC and seasoning encapsulated flavours.
 XX
 XX Sequence 2084 BP; 610 A; 401 C; 544 G; 529 T; 0 other;
 SO
 Query Match 80.0%; Score 16; DB 24; Length 2084;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 GGATGCTGGACACAA 18
 |||||
 DB 1292 GGATGCTGGACACAA 1277
 |||||
 RESULT 22
 AAS20026/c
 ID AAS20026 standard; DNA; 2084 BP.
 XX
 AC AAS20026;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Bacillus DNA encoding TERMAMYL-like alpha-amylase BAN.
 XX
 XX TERMAMYL; alpha-amylase; ds; detergent; dishwashing; textile desizing;
 KM starch liquefaction; ethanol production; hard surface cleaner;
 KM sweetener; amylopectin; limit dextrin; NOVAMYL; BAN.
 XX
 XX Bacillus amyloliquefaciens.
 OS
 XX Key Location/Qualifiers
 FH 343..1794
 FT CDS /*tag= a
 FT /*product= "Alpha-amylase BAN"
 FT /*partial
 FT /*note= "No start codon"
 FT
 XX WO200188107-A2.
 PN
 XX 22-NOV-2001.
 PD
 XX 10-MAY-2001; 2001WO-DK00323.
 PF
 XX 12-MAY-2000; 2000DK-0000779.
 PR
 XX (NOVO) NOVOZYMES AS.
 PA
 XX Svendsen A, Jorgensen CT, Nilsen BR;
 PI
 XX WPI; 2002-106123/14.
 DR
 XX P-PSDB; AAU2153.
 XX
 PT New variant of parent Termamyl-like alpha-amylase for use as a
 PT component in washing and dishwashing compositions, for textile

PT desizing, for starch liquefaction, and for producing sweeteners and
XX ethanol from starch -
XX Disclosure; Page 68-72; 84pp; English.
XX
XX The invention relates to a variant of parent TERMAMYL-like alpha-
CC amylase comprising an alteration at regions 186-193, 261-276, 283-293 or
CC 334-339, or at position 234, where the variant has alpha-amylase activity
CC and each position corresponds to a position of a parent Termamyl-like
CC alpha-amylase sequence having a Bacillus licheniformis alpha-amylase
CC sequence of 483 amino acids, given in specification. The variant alpha-
CC amylase, a detergent additive comprising the variant or a detergent
CC composition comprising the variant, is useful for washing and/or
CC dishwashing or textile desizing. The alpha-amylase is useful for starch
CC liquefaction or ethanol production and as a component in a hard surface
CC cleaning detergent composition, and for producing sweeteners from starch.
CC The variant has altered alpha-1, 6-D-glucosidic branch linkage
CC cleavage activity on amylopectin, preferably, increased alpha-1,
CC 6-D-glucosidic branch linkage cleavage activity of amylopectin or a
CC limit dextrin prepared by TERMAMYL (RTM) or NOVAMYL (RTM). The
CC present sequence encodes a natural variant of the TERMAMYL alpha-amylase,
CC BAN.
XX
SQ Sequence 2084 BP; 610 A; 401 C; 544 G; 529 T; 0 other;
Query Match 80.0%; Score 16; DB 24; Length 2084;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGGACACAA 18
DB 1292 GGATGCTGGACACAA 1277
RESULT 23
AAZ21079/c
ID AAZ21079 standard; DNA; 2604 BP.
XX
AC AAZ21079;
XX
DT 18-NOV-1999 (first entry)
XX
DE Bacillus amyloliquefaciens Termamyl-like alpha-amylase encoding DNA.
XX
XX Alpha-amylase; Termamyl-like alpha-amylase; glucose syrup; starch; ss.
XX
OS Bacillus amyloliquefaciens.
XX
XX Key Location/Qualifiers
FH -10_signal 707..712
FT /*tag= a
FT -35_signal 729..734
FT /*tag= b
FT RBS 759..762
FT /*tag= c
FT CDS 770..2314
FT /*tag= d
FT /*product= "Termamyl-like alpha-amylase"
FT sig_peptide 770..862
FT /*tag= e
FT mat_peptide 863..2311
FT /*tag= f
FT terminator 2321..2376
FT /*tag= g
XX
XX WO946399-A1.
XX
XX 16-SEP-1999.
XX
XX 08-MAR-1999; 99WO-DK00114.
XX
XX 09-MAR-1998; 98DK-0000321.
XX

PA (NOVO) NOVO-NORDISK AS.
XX
XX Norman BE, Hendriksen HV;
XX
XX WPI; 1999-551422/46.
XX
XX P-PSDB; AAY29853.
XX
XX Preparation of a glucose syrup, using a Termamyl-like alpha-amylase
PT
XX Disclosure; Page 29-32; 36pp; English.
XX
XX A method has been developed for the preparation of a glucose syrup using
CC Termamyl-like alpha-amylase containing a substitution at Val(54).
CC The glucose syrup obtained by the process is useful as an ingredient in
CC food products. The Termamyl-like alpha-amylase facilitates the
CC preparation of glucose syrups suitable for the food industry, previously
CC only possible using acid hydrolysis. The present sequence encodes
CC Bacillus amyloliquefaciens termamyl-like alpha-amylase.
XX
SQ Sequence 2604 BP; 755 A; 516 C; 684 G; 649 T; 0 other;
Query Match 80.0%; Score 16; DB 20; Length 2604;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGGACACAA 18
DB 1812 GGATGCTGGACACAA 1797
RESULT 24
AAA37850/c
ID AAA37850 standard; DNA; 2604 BP.
XX
AC AAA37850;
XX
DT 12-FEB-2001 (first entry)
XX
DE B. amyloliquefaciens termamyl-like alpha amylase coding sequence.
XX
XX Termamyl-like alpha-amylase; variant; starch liquefaction; fuel;
XX detergent composition; laundry cleaning composition; ethanol production;
XX dish washing cleaning composition; hard surface cleaning composition;
XX industrial ethanol production; textile desizing; ds.
XX
OS Bacillus amyloliquefaciens.
XX
XX Key Location/Qualifiers
FH -10_signal 707..712
FT /*tag= a
FT -35_signal 729..734
FT /*tag= b
FT RBS 759..762
FT /*tag= c
FT CDS 770..2314
FT /*tag= d
FT /*product= alpha-amylase
FT /note= "the signal peptide is not shown in the encoded
FT protein given in the specification"
FT sig_peptide 770..862
FT /*tag= e
FT mat_peptide 863..2311
FT /*tag= f
FT terminator 2321..2376
FT /*tag= g
XX
XX WO200060059-A2.
XX
XX 12-OCT-2000.
XX
XX 28-MAR-2000; 2000WO-DK00148.
XX

XX 30-MAR-1999; 99DK-0000437.
 PR (NOVO) NOVO NORDISK AS.
 XX
 PA
 XX
 PI Andersen C, Jorgensen CT, Bisgard-Frantzen H, Svendsen A;
 PI Kjaerulff S;
 XX WPI; 2001-015656/02.
 DR P-PSDB; AAY97546.
 XX
 PT New variants of parent Termamy1-like alpha-amylase, useful in starch
 PT liquefaction, in detergent compositions and in ethanol production,
 PT exhibit altered cleavage pattern relative to the parent -
 XX
 PS Disclosure; Page 64-67; 78pp; English.
 XX
 CC This sequence encodes a termamy1-like alpha amylase.
 CC The invention relates to a variant (I) of parent Termamy1-like
 CC alpha-amylase comprising alteration at one or more of the positions
 CC M13, G48, T49, S50, Q51, A52, D53, V54, G57, G107, G108, A111, S168 and
 CC M197. The alterations in (I) are independently an insertion of an amino
 CC acid downstream of the amino acid which occupies the position or deletion
 CC or substitution of the amino acid which occupies the position with a
 CC different amino acid. The variant has alpha-amylase activity. (I) or
 CC compositions containing it are useful in starch liquefaction, in
 CC detergent composition such as laundry, dish washing and hard surface
 CC cleaning compositions, ethanol production such as fuel, drinking and
 CC industrial ethanol production, desizing of textiles, fabrics or garments.
 CC (I) exhibits a reduced capability of cleaving a substrate closely to the
 CC branching point, and further exhibits improved substrate specificity
 CC and/or improved specific activity relative to the parent alpha-amylase.
 XX
 SQ Sequence 2604 BP; 755 A; 516 C; 684 G; 649 T; 0 other;
 XX
 Query Match 80.0%; Score 16; DB 22; Length 2604;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 3 GGATGCTGGACACAA 18
 Db 1812 GGATGCTGGACACAA 1797
 XX
 RESULT 25
 ID ABA47347/C
 XX ABA47347 standard; DNA; 364 BP.
 XX
 AC ABA47347;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #6042.
 XX
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-496933/54.
 XX
 PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 4; SEQ ID NO 6042; 327pp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BR 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;
 XX
 Query Match 75.0%; Score 15; DB 22; Length 364;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 5 ATGCTCGACACAG 19
 Db 76 ATGCTCGACACAG 62
 XX
 RESULT 26
 ID ABA65232/C
 XX ABA65232 standard; DNA; 364 BP.
 XX
 AC ABA65232;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #13537.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 KW
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX

```

XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver.
XX PS Claim 4; SEQ ID NO 13537; 639bp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;

Query Match      75.0%; Score 15; DB 22; Length 364;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      5 ATGCGCTGGACACCAAG 19
        |||||
        76 ATGCGCTGGACACCAAG 62

Db

RESULT 27
ABR32333/c
ID ABR32333 standard; DNA; 364 BP.
XX ABR32333;
XX AC ABR32333;
XX DT 23-JAN-2002 (first entry)
XX DE Probe #10799 for gene expression analysis in human heart cell sample.
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
XX OS Homo sapiens.
XX PN WO2001:57274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX CC Claim 4; SEQ ID NO 10799; 530bp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for

```

```

CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;

Query Match      75.0%; Score 15; DB 22; Length 364;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      5 ATGCGCTGGACACCAAG 19
        |||||
        76 ATGCGCTGGACACCAAG 62

Db

RESULT 28
AAK13649/c
ID AAK13649 standard; DNA; 364 BP.
XX AAK13649;
XX AC AAK13649;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe SEQ ID NO: 13640.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000667.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX CC Example 4; SEQ ID NO: 13640; 650bp + Sequence listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX SQ Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;

Query Match      75.0%; Score 15; DB 22; Length 364;

```

Best Local Similarity 100.0%; Pred. No. 29;
Matches 15; Conservat 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ATGCTTGACACAG 19
|||||
Db 76 ATGCTTGACACAG 62

RESULT 29
AAK39390/c

ID AAK39390 standard; DNA; 364 BP.

XX AAK39390;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 13947.

XX Human; bone marrow expressed exon; gene expression analysis; probe;
KM microarray; cancer; leukemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 13947; 658bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX bone marrow. They can be used to measure gene expression in bone marrow

XX samples, which may enable the improved diagnosis and treatment of cancers

XX such as lymphoma, leukemia and myeloma. The present sequence is one of

XX the probes of the invention.

XX Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;

XX Query Match 75.0%; Score 15; DB 22; Length 364;

XX Best Local Similarity 100.0%; Pred. No. 29;

XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ATGCTTGACACAG 19
|||||
Db 76 ATGCTTGACACAG 62

XX Probe #10136 for gene expression analysis in human cervical cell sample.
DE Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 25; SEQ ID NO 10136; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SENP). The present sequence is one such probe. The SENPs are derived

XX from human HeLa cells. The SENPs can be used to produce a single exon

XX sample derived from human cervical epithelial cells. By measuring gene

XX expression, the probes are therefore useful in grading and/or staging

XX of diseases of the cervix, notably cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIFO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;

XX Query Match 75.0%; Score 15; DB 22; Length 364;

XX Best Local Similarity 100.0%; Pred. No. 29;

XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ATGCTTGACACAG 19
|||||
Db 76 ATGCTTGACACAG 62

XX RESULT 31
AAI45403/c

XX ID AAI45403 standard; DNA; 364 BP.

XX AAI45403;

XX 17-OCT-2001 (first entry)

XX Probe #14089 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.
 PF 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488897/53. 2
 XX WPI; 2001-488897/53. 2
 DR Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta -
 PT
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 CC
 XX Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;
 SQ
 Query Match 75.0%; Score 15; DB 22; Length 364;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 5 ATGCTTGACACCAAG 19
 76 ATGCTTGACACCAAG 62
 QY
 Db
 RESULT 32
 AAI05907/c
 ID AAI05907 standard; DNA; 364 BP.
 AC AAI05907;
 XX
 DT 09-OCT-2001 (first entry)
 DE Probe #5898 used to measure gene expression in human breast sample.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 OS Homo sapiens.
 XX
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-476286/51.
 DR

XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 25; SEQ ID No 5898; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcr_sequences.
 CC
 XX Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;
 SQ
 Query Match 75.0%; Score 15; DB 22; Length 364;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 5 ATGCTTGACACCAAG 19
 76 ATGCTTGACACCAAG 62
 QY
 Db
 RESULT 33
 ABS38978/c
 ID ABS38978 standard; DNA; 364 BP.
 AC ABS38978;
 XX
 DT 25-FEB-2003 (first entry)
 DE Human liver single exon probe, SEQ ID No 13968.
 XX
 KW Human; single exon nucleic acid probe; liver; cirrhosis;
 KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
 KW coronary heart disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157273-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00664.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488898/53.
 XX WPI; 2001-488898/53.
 DR Human genome-derived single exon nucleic acid probes useful for
 XX analysing gene expression in human adult liver -
 PT
 CC Claim 4; SEQ ID No 13968; 658pp; English.
 XX
 CC The invention relates to a single exon nucleic acid probe (SENP) (I) for

CC measuring human gene expression in a sample derived from human adult
 CC liver, comprising one of 1109 defined nucleotide sequences given in the
 CC specification (or complements/ fragments). The probe hybridises at high
 CC stringency to a nucleic acid molecule expressed in the human adult
 CC liver. (1) may be used for predicting, measuring and displaying gene
 CC expression in samples derived from human adult liver. The genes
 CC identified may be involved in genetic liver diseases such as cirrhosis,
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
 CC is associated with coronary heart disease. AB525011-AB551005 represent
 CC human liver single exon nucleic acid probes of the invention.
 CC Note: The sequence information for this patent does not appear in the
 CC printed specification but was obtained in electronic format directly
 CC from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
 CC
 CC Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;
 CC
 CC Query Match 75.0%; Score 15; DB 23; Length 364;
 CC Best Local Similarity 100.0%; Pred. No. 29;
 CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC 5 ATGCCTGGACACAG 19
 CC 76 ATGCCTGGACACAG 62
 CC
 CC RESULT 34
 CC ABS13477/c
 CC ID ABS13477 standard; DNA; 364 BP.
 CC
 CC AC ABS13477;
 CC XX
 CC DT 19-AUG-2002 (first entry)
 CC
 CC DE Human genome-derived single exon probe ORF from lung SEQ ID No 13468.
 CC
 CC XX Human; ds; single exon probe; asthma; lung cancer; COPD; IHD;
 CC XX chronic obstructive pulmonary disease; interstitial lung disease;
 CC XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
 CC XX tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;
 CC XX Hereditary-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
 CC XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;
 CC XX primary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 CC XX primary ciliary dyskinesia; pulmonary hypertension;
 CC XX hyaline membrane disease; open reading frame; ORF.
 CC
 CC OS Homo sapiens.
 CC
 CC XX MO200186003-A2.
 CC XX
 CC PD 15-NOV-2001.
 CC
 CC PF 30-JAN-2001; 2001WO-U00665.
 CC
 CC PR 04-FEB-2000; 2000US-180312P.
 CC PR 26-MAY-2000; 2000US-207456P.
 CC PR 30-JUN-2000; 2000US-0608409.
 CC PR 03-AUG-2000; 2000US-063236P.
 CC PR 21-SEP-2000; 2000US-234687P.
 CC PR 27-SEP-2000; 2000US-236359P.
 CC PR 04-OCT-2000; 2000GB-0024263.
 CC
 CC PA (MOLE-) MOLECULAR DYNAMICS INC.
 CC XX Penn SG, Hanzel DK, Chen W, Rank DR;
 CC XX
 CC DR WPI; 2002-114183/15.
 CC
 CC XX Spatially-addressable set of single exon nucleic acid probes, used to
 CC PT measure gene expression in human lung samples -
 CC PS Claim 4; SEQ ID No 13468, 6346P; English.
 CC XX
 CC CC The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe;
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 1201 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer; chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberosus sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hereditary-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
 CC
 CC SQ Sequence 364 BP; 94 A; 94 C; 75 G; 101 T; 0 other;
 CC
 CC Query Match 75.0%; Score 15; DB 24; Length 364;
 CC Best Local Similarity 100.0%; Pred. No. 29;
 CC Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 5 ATGCCTGGACACAG 19
 CC DB 76 ATGCCTGGACACAG 62
 CC
 CC RESULT 35
 CC ABA36167/c
 CC ID ABA36167 standard; DNA; 393 BP.
 CC
 CC AC ABA36167;
 CC XX
 CC DT 23-JAN-2002 (first entry)
 CC
 CC DE Probe #14633 for gene expression analysis in human heart cell sample.
 CC
 CC XX Human; gene expression; heart; microarray; vascular system; probe;
 CC XX cardiovascular disease; hypertension; cardiac arrhythmia;
 CC XX congenital heart disease; ss.
 CC
 CC OS Homo sapiens.
 CC
 CC XX MO200157274-A2.
 CC XX
 CC PD 09-AUG-2001.
 CC
 CC PF 30-JAN-2001; 2001WO-US00666.
 CC XX

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0924263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-48899/53.
 XX
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 XX
 PS Claim 4; SEQ ID No 14633; 530bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 393 BP; 104 A; 100 C; 82 G; 107 T; 0 other;
 Query Match 75.0%; Score 15; DB 22; Length 393;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 ATGCGTGACACAG 19
 Db 85 ATGCGTGACACAG 71

Search completed: August 14, 2003, 21:41:27
 Job time : 132 secs

101 14 63.6 240 3 US-09-559-397A-9 Sequence 9, Appl
102 14 63.6 258 4 US-09-252-991A-16026 Sequence 16026, A
103 14 63.6 261 4 US-09-313-284A-3461 Sequence 3461, Ap
104 14 63.6 274 4 US-09-313-284A-1061 Sequence 1061, Ap
105 14 63.6 356 3 US-08-896-164-17 Sequence 17, Appl
106 14 63.6 357 4 US-09-252-991A-4619 Sequence 4619, Ap
107 14 63.6 361 4 US-09-386-642-1 Sequence 1, Appl
108 14 63.6 378 4 US-09-252-991A-4519 Sequence 3474, Ap
109 14 63.6 378 4 US-09-252-991A-4519 Sequence 4519, Ap
110 14 63.6 382 4 US-09-386-642-4 Sequence 4, Appl
111 14 63.6 417 4 US-09-252-991A-3560 Sequence 3560, Ap
112 14 63.6 468 4 US-09-252-991A-16533 Sequence 16533, A
113 14 63.6 474 4 US-09-252-991A-5945 Sequence 5945, Ap
114 14 63.6 484 4 US-09-386-642-3 Sequence 3, Appl
115 14 63.6 492 4 US-09-199-637A-334 Sequence 334, Ap
116 14 63.6 510 4 US-09-252-991A-6552 Sequence 6552, Ap
117 14 63.6 532 3 US-09-060-756-181 Sequence 181, Ap
118 14 63.6 532 4 US-09-670-314-181 Sequence 181, Ap
119 14 63.6 576 4 US-09-252-991A-7453 Sequence 7453, Ap
120 14 63.6 621 4 US-09-252-991A-4261 Sequence 4261, Ap

ALIGNMENTS

RESULT 1
US-09-311-260-84/C
Sequence 84, Application US/09311260
Patent No. 6214555
GENERAL INFORMATION:
APPLICANT: Ieushner, James
APPLICANT: Hull, May
APPLICANT: Lacroix, Jean-Michel
TITLE OF INVENTION: METHOD, COMPOSITIONS AND KIT FOR DETECTION OF
TITLE OF INVENTION: MICROORGANISMS AND BI-DIRECTIONAL SEQUENCING OF NUCLEIC ACID
TITLE OF INVENTION: POLYMERS
NUMBER OF SEQUENCES: 189
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oppedahl & Larson LLP
STREET: P. O. Box 5270
CITY: Frisco
STATE: CO
COUNTRY: US
ZIP: 80443-5270
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb storage
COMPUTER: IBM compatible
OPERATING SYSTEM: MS DOS
SOFTWARE: Word Perfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/311,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Larson, Marina T.
REGISTRATION NUMBER: 32,038
REFERENCE/DOCKET NUMBER: VGEN-P-058-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (970) 668-2050
TELEFAX: (970) 668-2082
TELEX:
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 22
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
HYPOTHETICAL: no

ANTI-SENSE: no
FRAGMENT TYPE: internal
US-09-311-260-84

Query Match 100.0%; Score 22; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 22 ATGCCACACGACGACCA 22
22 ATGCCACACGACGACCA 1

RESULT 2
US-09-252-991A-14225
Sequence 14225, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 14225

LENGTH: 549
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14225

Query Match 81.8%; Score 18; DB 4; Length 549;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 5 CCACGACGACGACCA 22
328 CCACGACGACGACCA 345

RESULT 3
US-09-252-991A-14137
Sequence 14137, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 14137
LENGTH: 696
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14137

Query Match 81.8%; Score 18; DB 4; Length 696;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 5 CCACGACGACGACCA 22
423 CCACGACGACGACCA 440

Mon Aug 18 10:30:11 2003

us-10-074-620-6.011.m1

Page 3

```
RESULT 4
US-08-961-527-12
; Sequence 12, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunesh
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-12

Query Match      81.8%; Score 18; DB 4; Length 9909;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 CCACCACGACGACGACCA 22
Db      7540 CCACCACGACGACGACCA 7557

RESULT 5
US-09-252-991A-119/c
; Sequence 119, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 119
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
```

```
US-09-252-991A-119

Query Match      77.3%; Score 17; DB 4; Length 417;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 CACCACGACGACGACCA 22
Db      105 CACCACGACGACGACCA 89

RESULT 6
US-09-252-991A-10582/c
; Sequence 10582, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10582
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-10582

Query Match      77.3%; Score 17; DB 4; Length 615;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 GCCACGACGACGACGAC 20
Db      348 GCCACGACGACGACGAC 332

RESULT 7
US-09-252-991A-132/c
; Sequence 132, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 132
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-132

Query Match      77.3%; Score 17; DB 4; Length 717;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 CACCACGACGACGACCA 22
Db      91 CACCACGACGACGACCA 75
```

RESULT 8
US-08-911-853-24/c
Sequence 24, Application US/08911853
Patent No. 6048710
GENERAL INFORMATION:
APPLICANT: Gerritse, Gijssbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,853
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,092
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-911-853-24
Query Match 77.3%; Score 17; DB 3; Length 954;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 GCCACCACCGACGAC 20
DB 50 GCCACCACCGACGAC 34
RESULT 9
US-09-479-409-24/c
Sequence 24, Application US/09479409
Patent No. 6225106
GENERAL INFORMATION:
APPLICANT: Gerritse, Gijssbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/479,409
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-479-409-24
Query Match 77.3%; Score 17; DB 3; Length 954;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 GCCACCACCGACGAC 20
DB 50 GCCACCACCGACGAC 34
RESULT 10
US-09-479-453-24/c
Sequence 24, Application US/09479453
Patent No. 6313283
GENERAL INFORMATION:
APPLICANT: Gerritse, Gijssbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/479,453
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-846-7620
TELEFAX: 650-845-6504
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-479-453-24

Query Match 77.3%; Score 17; DB 4; Length 954;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GCCACCCAGCAGCAGC 20
Db 50 GCCACCCAGCAGCAGC 34

RESULT 11
US-09-252-991A-10275
Sequence 10275; Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 10275
LENGTH: 984
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-10275 2

Query Match 77.3%; Score 17; DB 4; Length 984;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GCCACCCAGCAGCAGC 20
Db 865 GCCACCCAGCAGCAGC 881

RESULT 12
US-09-252-991A-146
Sequence 146; Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 146
LENGTH: 1035
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa

Query Match 77.3%; Score 17; DB 4; Length 1035;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 CACCACCCAGCAGCAGC 22
Db 234 CACCACCCAGCAGCAGC 250

RESULT 13

US-09-252-991A-12112/c
Sequence 12112; Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 12112
LENGTH: 1485
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12112

Query Match 77.3%; Score 17; DB 4; Length 1485;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GCCACCCAGCAGCAGC 20
Db 536 GCCACCCAGCAGCAGC 520

RESULT 14
US-09-252-991A-10486
Sequence 10486; Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 10486
LENGTH: 2778
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa

Query Match 77.3%; Score 17; DB 4; Length 2778;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GCCACCCAGCAGCAGC 20
Db 1103 GCCACCCAGCAGCAGC 1119

RESULT 15
US-09-548-938A-2
Sequence 2; Application US/09548938A
Patent No. 6573086
GENERAL INFORMATION:
APPLICANT: EMALFARB, MARK AARON
APPLICANT: BURLINGAME, RICHARD PAUL
APPLICANT: OLSON, PHILIP TERRY
APPLICANT: SINITSYN, ARIADY PANTELEIMONOVICH
APPLICANT: PARICHE, MARTINE
APPLICANT: BOUSSON, JEAN CHRISTOPHE
APPLICANT: PYNMONEN, CHRISTINE MARIE

APPLICANT: PUNT, PETER JAN
TITLE OF INVENTION: TRANSPORTATION SYSTEM IN THE FIELD OF FILAMENTOUS FUNGI
FILE REFERENCE: 3123-4001
CURRENT APPLICATION NUMBER: US/09/548,938A
CURRENT FILING DATE: 2000-04-13
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 2
LENGTH: 3028
TYPE: DNA
ORGANISM: Chrysoosporium lucknowense
FEATURE:
NAME/KEY: modified base
LOCATION: (46)...(47)
OTHER INFORMATION: a, t, c, g, other or unknown
FEATURE:
NAME/KEY: modified base
LOCATION: (374)
OTHER INFORMATION: a, t, c, g, other or unknown
US-09-548-938A-2

Query Match 77.3%; Score 17; DB 4; Length 3028;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 CACACACACACACACCA 22
DB 1380 CACACACACACACCA 1396

RESULT 16
US-09-252-991A-10685/C
Sequence 10685, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 10685
LENGTH: 3339
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10685

Query Match 77.3%; Score 17; DB 4; Length 3339;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACACACACACACAC 20
DB 1661 GCCACACACACACACAC 1645

RESULT 17
US-08-790-374-1/C
Sequence 1, Application US/08790374
Patent No. 5863734
GENERAL INFORMATION:
APPLICANT: Karavayogu, Maria
TITLE OF INVENTION: METHODS OF TREATMENT FOR OBSESSIVE-COMPULSIVE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson

D

STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/790,374
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-189
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3651 base pairs
TYPE: nucleic acid
STANDARDS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
US-08-790-374-1

Query Match 77.3%; Score 17; DB 2; Length 3651;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 CACACACACACACACCA 22
DB 1403 CACACACACACACCA 1387

RESULT 18
US-08-911-853-29/C
Sequence 29, Application US/08911853
Patent No. 6048710
GENERAL INFORMATION:
APPLICANT: Geritise, Gijssert
TITLE OF INVENTION: QUAX, WILHELMUS J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,853
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,092
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gialster, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-846-7620
 TELEFAX: 650-846-6504
 INFORMATION FOR SEQ ID NO: 29:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 17612 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-08-911-853-29

Query Match 77.3%; Score 17; DB 3; Length 17612;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACGAC 20
 DB 14918 GCCACCACGACGACGAC 14902

RESULT 19
 US-09-479-409-29/c

Sequence 29, Application US/09479409
 Patent No. 6225106
 GENERAL INFORMATION:
 APPLICANT: Gerritse, Gijbert
 APPLICANT: Quax, Wilhelmus J.
 TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
 TITLE OF INVENTION: EXPRESSION LEVELS
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genencor International
 STREET: 925 Page Mill Road
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1013
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/479,409
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/911,853
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Glaister, Debra J.
 REGISTRATION NUMBER: 33,888
 REFERENCE/DOCKET NUMBER: GC361-2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-846-7620
 TELEFAX: 650-846-6504
 INFORMATION FOR SEQ ID NO: 29:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 17612 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-09-479-409-29

Query Match 77.3%; Score 17; DB 3; Length 17612;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACGAC 20
 DB 14918 GCCACCACGACGACGAC 14902

RESULT 20

US-09-479-453-29/c

Sequence 29, Application US/09479453
 Patent No. 6313283
 GENERAL INFORMATION:
 APPLICANT: Gerritse, Gijbert
 APPLICANT: Quax, Wilhelmus J.
 TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
 TITLE OF INVENTION: EXPRESSION LEVELS
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genencor International
 STREET: 925 Page Mill Road
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1013
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/479,453
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/911,853
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Glaister, Debra J.
 REGISTRATION NUMBER: 33,888
 REFERENCE/DOCKET NUMBER: GC361-2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-846-7620
 TELEFAX: 650-846-6504
 INFORMATION FOR SEQ ID NO: 29:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 17612 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

US-09-479-453-29

Query Match 77.3%; Score 17; DB 4; Length 17612;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCACGACGACGAC 20
 DB 14918 GCCACCACGACGACGAC 14902

RESULT 21

US-08-990-823-48/c
 Sequence 48, Application US/08990823D
 Patent No. 6228371

GENERAL INFORMATION:
 APPLICANT: Nano, Francis
 TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 FILE REFERENCE: 49086
 CURRENT APPLICATION NUMBER: US/08/990,823D
 CURRENT FILING DATE: 1997-12-15
 EARLIER APPLICATION NUMBER: US 96/10375
 EARLIER FILING DATE: 1996-06-14
 EARLIER APPLICATION NUMBER: 60/000,254
 EARLIER FILING DATE: 1995-06-15
 NUMBER OF SEQ ID NOS: 113
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 48
 LENGTH: 464
 TYPE: DNA
 ORGANISM: Mycobacterium tuberculosis
 FEATURE:

NAME/KEY: Modified base
OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-48

Query Match 72.7%; Score 16; DB 3; Length 464;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ACCACGACGACGACCA 22
DB 221 ACCACGACGACGACCA 206

RESULT 22
US-09-477-135A-48/c
Sequence 48, Application US/09477135A
Patent No. 6572865
GENERAL INFORMATION:
APPLICANT: Nemo, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
FILE REFERENCE: 52888
CURRENT APPLICATION NUMBER: US/09/477,135A
PRIOR FILING DATE: 2000-01-03
PRIOR APPLICATION NUMBER: 08990823
PRIOR FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: US 96/10375
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 48
LENGTH: 464
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: variation
LOCATION: (1)..(464)
OTHER INFORMATION: n = A, C, G, or T
US-09-477-135A-48

Query Match 72.7%; Score 16; DB 4; Length 464;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ACCACGACGACGACCA 22
DB 221 ACCACGACGACGACCA 206

RESULT 23
US-09-252-991A-2258
Sequence 2258, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196,136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 2258
LENGTH: 951
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2258

Query Match 72.7%; Score 16; DB 4; Length 951;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACGACGACGACGAC 20
DB 545 CCACGACGACGACGAC 560

RESULT 24
US-09-252-991A-7899/c
Sequence 7899, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196,136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 7899
LENGTH: 966
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7899

Query Match 72.7%; Score 16; DB 4; Length 966;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACGACGACGACCA 19
DB 416 GCCACGACGACGACCA 401

RESULT 25
US-09-252-991A-16455/c
Sequence 16455, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196,136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16455
LENGTH: 966
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16455

Query Match 72.7%; Score 16; DB 4; Length 966;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACGACGACGACGAC 20
DB 367 CCACGACGACGACGAC 352

RESULT 26
US-09-252-991A-16148

Sequence 16148, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16148
LENGTH: 1116
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16148

Query Match 72.7%; Score 16; DB 4; Length 1116;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACCACCGACGAC 20
DB 765 CCACCACCGACGAC 780

RESULT 27
US-08-976-259-74
Sequence 74, Application US/08976259
Patent No. 631609
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Choi, Gail H.
APPLICANT: Welch, Rodney A.
TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
Patent No. 631609
NUMBER OF SEQUENCES: 142
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Ave, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,259
FILING DATE: Herewith
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488-0740002/EKS/CM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2640
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1332 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-976-259-74

Query Match 72.7%; Score 16; DB 4; Length 1332;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTGCCACCGACGACG 17
DB 283 TTGCCACCGACGACG 298

RESULT 28
US-09-252-991A-2505/C
Sequence 2505, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 2505
LENGTH: 1512
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2505

Query Match 72.7%; Score 16; DB 4; Length 1512;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACCACCGACGAC 20
DB 1054 CCACCACCGACGAC 1039

RESULT 29
US-08-365-981-5/C
Sequence 5, Application US/08365981
Patent No. 5583030
GENERAL INFORMATION:
APPLICANT: Robert DICKSON et al.
TITLE OF INVENTION: METHOD FOR OBTAINING
TITLE OF INVENTION: ANTIFUNGAL AND HERBICIDAL COMPOUNDS THAT TARGET THE
TITLE OF INVENTION: FIRST COMMITTED STEP IN SPHINGOLIPID LONG-CHAIN
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: LOWE, PRICE, LEBLANC & BECKER
STREET: STE. 300, 99 CANAL CENTER PLAZA
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: DOS Text file
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/365,981
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/906,899
FILING DATE: 06/30/92
ATTORNEY/AGENT INFORMATION:
NAME: D.J. Mills
REGISTRATION NUMBER: 34506

REFERENCE/DOCKET NUMBER: 434-028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 684-1111
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1683
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: polypeptide
US-08-365-981-5

Query Match 72.7%; Score 16; DB 1; Length 1683;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ACCACGACGACGACCA 22
DB 1118 ACCACGACGACGACCA 1103

RESULT 30
US-09-252-991A-2424/c
Sequence 2424, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 2424
LENGTH: 1788
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2424

Query Match 72.7%; Score 16; DB 4; Length 1788;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACGACGACGACGAC 20
DB 980 CCACGACGACGACGAC 965

RESULT 31
US-09-252-991A-16341/c
Sequence 16341, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16341
LENGTH: 2019
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16341

Query Match 72.7%; Score 16; DB 4; Length 2019;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACGACGACGACGAC 20
DB 245 CCACGACGACGACGAC 230

RESULT 32
US-09-252-991A-7660
Sequence 7660, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 7660
LENGTH: 2946
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7660

Query Match 72.7%; Score 16; DB 4; Length 2946;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACGACGACGACCA 19
DB 524 GCCACGACGACGACCA 539

RESULT 33
US-08-475-035-3/c
Sequence 3, Application US/08475035
Patent No. 5985553
GENERAL INFORMATION:
APPLICANT: KING, C. R.
APPLICANT: KRAUS, MATTHIAS H.
APPLICANT: AARONSON, STUART A.
TITLE OF INVENTION: HUMAN GENE RELATED TO BUT DISTINCT FROM
TITLE OF INVENTION: EGF RECEPTOR GENE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: NEEDLE & ROSENBERG, P.C.
STREET: Suite 1200, 127 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,035
FILING DATE: 7 Jun 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Perryman, David G.
REGISTRATION NUMBER: 33,438
REFERENCE/DOCKET NUMBER: 1414.656
TELECOMMUNICATION INFORMATION:

TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5532 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 187..3816
US-08-475-035-3

Query Match #2.7%; Score 16; DB 2; Length 5532;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACACCGACGCA 19
|||||
DB 2168 GCCACACCGACGCA 2153

RESULT 34
US-09-676-610B-17/c
Sequence 17, Application US/09676610B
Patent No. 6444465
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Jacqueline Wyatt
APPLICANT: Susan M. Freiler
TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
FILE REFERENCE: RTS-0138
CURRENT APPLICATION NUMBER: US/09/676,610B
CURRENT FILING DATE: 2000-09-29
NUMBER OF SEQ ID NOS: 182
SEQ ID NO 17
LENGTH: 5532
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (187) ... (3819)
US-09-676-610B-17

Query Match 72.7%; Score 16; DB 4; Length 5532;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACACCGACGCA 19
|||||
DB 2168 GCCACACCGACGCA 2153

RESULT 35
US-09-676-610B-24/c
Sequence 24, Application US/09676610B
Patent No. 6444465
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Jacqueline Wyatt
APPLICANT: Susan M. Freiler
TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
FILE REFERENCE: RTS-0138
CURRENT APPLICATION NUMBER: US/09/676,610B
CURRENT FILING DATE: 2000-09-29
NUMBER OF SEQ ID NOS: 182
SEQ ID NO 24
LENGTH: 169998
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: exon

LOCATION: (1208) ... (1472)
NAME/KEY: intron
LOCATION: (1473) ... (124390)
NAME/KEY: exon
LOCATION: (124391) ... (124544)
NAME/KEY: intron
LOCATION: (124545) ... (125409)
NAME/KEY: exon
LOCATION: (125410) ... (125595)
NAME/KEY: intron
LOCATION: (125596) ... (128711)
NAME/KEY: exon
LOCATION: (128712) ... (128848)
NAME/KEY: intron
LOCATION: (128849) ... (133400)
NAME/KEY: exon
LOCATION: (133401) ... (133469)
NAME/KEY: intron
LOCATION: (133470) ... (134652)
NAME/KEY: exon
LOCATION: (134653) ... (134773)
NAME/KEY: intron
LOCATION: (134774) ... (136116)
NAME/KEY: exon
LOCATION: (136117) ... (136261)
NAME/KEY: intron
LOCATION: (136262) ... (137936)
NAME/KEY: exon
LOCATION: (137937) ... (138053)
NAME/KEY: intron
LOCATION: (138054) ... (138637)
NAME/KEY: exon
LOCATION: (138638) ... (138766)
NAME/KEY: intron
LOCATION: (138767) ... (138864)
NAME/KEY: exon
LOCATION: (138865) ... (138940)
NAME/KEY: intron
LOCATION: (138941) ... (139765)
NAME/KEY: exon
LOCATION: (139766) ... (139860)
NAME/KEY: intron
LOCATION: (139861) ... (142245)
NAME/KEY: exon
LOCATION: (142246) ... (142445)
NAME/KEY: intron
LOCATION: (142446) ... (143605)
NAME/KEY: exon
LOCATION: (143606) ... (143738)
NAME/KEY: intron
LOCATION: (143739) ... (145838)
NAME/KEY: exon
LOCATION: (145839) ... (145931)
NAME/KEY: intron
LOCATION: (145932) ... (147385)
NAME/KEY: exon
LOCATION: (147386) ... (147544)
NAME/KEY: intron
LOCATION: (147545) ... (153274)
NAME/KEY: exon
LOCATION: (153275) ... (153321)
NAME/KEY: intron
LOCATION: (153322) ... (155088)
NAME/KEY: exon
LOCATION: (155089) ... (155231)
NAME/KEY: intron
LOCATION: (155232) ... (156025)
NAME/KEY: exon
LOCATION: (156026) ... (156151)
NAME/KEY: intron
LOCATION: (156152) ... (156826)
NAME/KEY: exon
LOCATION: (156827) ... (156928)

Mon Aug 18 10:30:11 2003

us-10-074-620-6.011.rn1

Page 12

; NAME/KEY: intron
; LOCATION: (156929)...(163399)
; NAME/KEY: exon
; LOCATION: (163400)...(163586)
US-09-676-610B-24

Query Match 72.7%; Score 16; DB 4; Length 163998;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCCACGACGCA 19
|||
Db 155152 GCCACCCACGACGCA 155137

Search completed: August 15, 2003, 11:00:47
Job time : 48.475 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:41:37 ; Search time 1252 Seconds

(without alignments)
388.250 Million cell updates/sec

Title: US-10-074-620-5

Perfect score: 20

Sequence: 1 agggatgctcgacacaga 20

Scoring table: OLIGO NUC

Searched: 22781392 seqs, 12152236056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 120 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hiv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pig:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	95.0	305	28	BR289461 CH330-158
2	19	95.0	451	14	CB787567 AMGNNUC:N
3	19	95.0	510	14	CB782526 AMGNNUC:N
4	19	95.0	518	14	CB715916 AMGNNUC:N

5	17	85.0	251	10	BB084267	BB084267
6	17	85.0	400	13	BY649980	BY649980
7	17	85.0	433	13	BY640733	BY640733
8	16	80.0	254	10	BP931900	BP931900
9	16	80.0	259	28	AO552781	AO552781
10	16	80.0	351	9	AV733084	AV733084
11	16	80.0	332	14	CB74635	CB74635
12	16	80.0	638	14	CB436177	CB436177
13	16	80.0	668	29	AG148078	AG148078
14	16	80.0	785	10	BE544293	BE544293
15	15	80.0	890	14	CA584064	CA584064
16	15	75.0	301	9	AM14383	AM14383
17	15	75.0	304	10	BB281631	BB281631
18	15	75.0	333	13	BQ328868	BQ328868
19	15	75.0	366	13	BQ328810	BQ328810
20	15	75.0	368	13	BY608974	BY608974
21	15	75.0	420	13	BQ490402	BQ490402
22	15	75.0	440	28	B49246	B49246
23	15	75.0	446	9	AL641014	AL641014
24	15	75.0	451	14	CB787356	CB787356
25	15	75.0	520	28	AQ359962	AQ359962
26	15	75.0	522	9	AM651297	AM651297
27	15	75.0	531	9	AU101494	AU101494
28	15	75.0	535	13	BX282004	BX282004
29	15	75.0	561	10	BB641771	BB641771
30	15	75.0	603	13	BQ489926	BQ489926
31	15	75.0	603	13	BQ591046	BQ591046
32	15	75.0	624	28	AQ157654	AQ157654
33	15	75.0	629	12	BM725133	BM725133
34	15	75.0	632	14	CB918954	CB918954
35	15	75.0	642	14	CB919968	CB919968
36	15	75.0	648	14	BB260326	BB260326
37	15	75.0	656	10	BB260326	BB260326
38	15	75.0	656	28	BZ131723	BZ131723
39	15	75.0	656	29	AG105990	AG105990
40	15	75.0	665	10	BG482911	BG482911
41	15	75.0	671	12	B1820618	B1820618
42	15	75.0	682	14	CB007655	CB007655
43	15	75.0	682	29	AG094204	AG094204
44	15	75.0	688	12	B1822450	B1822450
45	15	75.0	687	12	B1824619	B1824619
46	15	75.0	697	12	B1823937	B1823937
47	15	75.0	702	29	AG013830	AG013830
48	15	75.0	715	29	AG013852	AG013852
49	15	75.0	718	29	AG013833	AG013833
50	15	75.0	725	29	AG013831	AG013831
51	15	75.0	731	29	AG013851	AG013851
52	15	75.0	739	12	B1920655	B1920655
53	15	75.0	760	12	B1770406	B1770406
54	15	75.0	766	14	CD521744	CD521744
55	15	75.0	767	12	B1489632	B1489632
56	15	75.0	793	29	AG119349	AG119349
57	15	75.0	796	13	BU486319	BU486319
58	15	75.0	802	13	BU383029	BU383029
59	15	75.0	805	10	BG024904	BG024904
60	15	75.0	809	12	B1820265	B1820265
61	15	75.0	810	10	BF038235	BF038235
62	15	75.0	820	12	B1821733	B1821733
63	15	75.0	826	10	BG473154	BG473154
64	15	75.0	835	10	BG542824	BG542824
65	15	75.0	851	14	CB683410	CB683410
66	15	75.0	869	14	BP234305	BP234305
67	15	75.0	876	29	CN802050	CN802050
68	15	75.0	879	28	BZ172688	BZ172688
69	15	75.0	926	12	B1819407	B1819407
70	15	75.0	942	10	BF237645	BF237645
71	15	75.0	945	13	BQ682507	BQ682507
72	15	75.0	954	10	BG254521	BG254521
73	15	75.0	969	10	BG482095	BG482095
74	15	75.0	1041	29	CN803079	CN803079
75	15	75.0	1261	28	AQ738879	AQ738879
76	15	75.0	1373	11	AK043901	AK043901
77	15	75.0	1441	13	BQ225276	BQ225276

C	78	15	75.0	4204	11	AK047115
C	79	14	70.0	558	9	AI900733
C	80	14	70.0	155	9	AV414569
C	81	14	70.0	211	9	AI100937
C	82	14	70.0	220	10	BG372712
C	83	14	70.0	227	10	BG037234
C	84	14	70.0	250	9	AV427989
C	85	14	70.0	251	29	BZ746090
C	86	14	70.0	263	28	AZ562270
C	87	14	70.0	273	10	BF457179
C	88	14	70.0	274	28	AZ025461
C	89	14	70.0	277	9	AI638823
C	90	14	70.0	277	9	AI655052
C	91	14	70.0	277	9	AI671554
C	92	14	70.0	285	13	BQ325498
C	93	14	70.0	287	28	BZ096268
C	94	14	70.0	288	9	AI278622
C	95	14	70.0	288	13	BQ813572
C	96	14	70.0	290	10	BF872835
C	97	14	70.0	292	10	BB243581
C	98	14	70.0	305	9	AV167330
C	99	14	70.0	307	10	BB121635
C	100	14	70.0	309	9	AI433067
C	101	14	70.0	313	9	AA357110
C	102	14	70.0	314	10	BB342912
C	103	14	70.0	317	14	CB708753
C	104	14	70.0	321	9	AA982412
C	105	14	70.0	324	9	AI841944
C	106	14	70.0	324	29	CNS03KGB
C	107	14	70.0	325	10	BE122542
C	108	14	70.0	328	12	BM766562
C	109	14	70.0	330	12	BM702585
C	110	14	70.0	331	9	AI011674
C	111	14	70.0	331	9	AI318455
C	112	14	70.0	334	12	BM665296
C	113	14	70.0	337	12	BM743219
C	114	14	70.0	337	28	AQ110626
C	115	14	70.0	343	10	BB241726
C	116	14	70.0	345	14	M13662
C	117	14	70.0	350	29	BZ318794
C	118	14	70.0	356	12	BM651541
C	119	14	70.0	358	10	BG057262
C	120	14	70.0	358	10	BE096603

ALIGNMENTS

RESULT 1
BH289461/c
LOCUS
DEFINITION CH230-158G22.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-158G22, genomic survey sequence.

ACCESSION
BH289461
VERSION
BH289461.1
KEYWORDS
GSS.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 305)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
,A., Gebregiorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhac@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rac230.html). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_oring_information.htm). BAC end
pages: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 158 row: G column: 22
Seq primer: SP6
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..305
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SNHed/MCM"
/db_xref="taxon:10116"
/clone="CH230-158G22"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/note="Vector: pTARbac2.1; Site_1: EcoRI; Site_2: EcoRI;
CHORI-230 Rat (BN/SNHed/MCM) BAC library produced by
Pieter de Jong"

BASE COUNT
74 a 100 c 56 g 75 t

Query Match
Best Local Similarity 95.0%; Score 19; DB 28; Length 305;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 AGGATGCTGACACAG 19
249 AGGATGCTGACACAG 231

RESULT 2
CB787567
LOCUS
DEFINITION AMGNIC.NRW3-00062-B3-A white adipose tise (10469) Rattus
norvegicus cDNA clone nrw3-00062-b3 5', mRNA sequence.

ACCESSION
CB787567
VERSION
CB787567.1
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 451)
Amgen EST Program.
Amgen Rat EST Program
Unpublished
Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00062 row: b column: 3.

FEATURES
source
Location/Qualifiers
1..451
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrw3-00062-b3"
/tissue_type="adipose tise"
/clone_lib="white adipose tise (10469)"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; white
adipose tise adult female Wistar rats; >250 grams"

BASE COUNT
75 a 97 c 151 g 67 t 61 others

Query Match
Best Local Similarity 95.0%; Score 19; DB 14; Length 451;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGACACAG 19
 |||||
 DB 219 AGGATGCTGACACAG 237

RESULT 3
 CB782526 510 bp mRNA linear EST 16-MAY-2003
 LOCUS AMGNNUC:NRDGI-00179-D6-A nrdg1 (10855) Rattus norvegicus cDNA clone
 DEFINITION nrdg1-00179-d6 5', mRNA sequence.
 ACCESSION CB782526
 VERSION CB782526.1 GI:29870917
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (bases 1 to 510)
 AUTHORS Amgen EST Program.
 TITLE Amgen Rat EST Program
 JOURNAL Unpublished
 COMMENT Contact: Dan Fitzpatrick
 Amgen, Inc
 One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
 Tel: 805 447-4881
 Plate: 00179 row: d column: 6.
 Location/Qualifiers
 1..510
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="nrdg1-00179-d6"
 /cissue_type="Dorsal Root Ganglia"
 /clone_lib="nrdg1 (10855)"
 /note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; rat
 dorsal root ganglia"
 95 a 143 c 191 g 81 t

BASE COUNT 95 a 143 c 191 g 81 t

ORIGIN

Query Match 95.0%; Score 19; DB 14; Length 510;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGACACAG 19
 |||||
 DB 145 AGGATGCTGACACAG 163

RESULT 4
 CB715916 518 bp mRNA linear EST 10-APR-2003
 LOCUS AMGNNUC:NRDGI-00190-D9-A nrdg1 (10855) Rattus norvegicus cDNA clone
 DEFINITION nrdg1-00190-d9 5', mRNA sequence.
 ACCESSION CB715916
 VERSION CB715916.1 GI:29773064
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (bases 1 to 518)
 AUTHORS Amgen EST Program.
 TITLE Amgen Rat EST Program
 JOURNAL Unpublished
 COMMENT Contact: Dan Fitzpatrick
 Amgen, Inc
 One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
 Tel: 805 447-4881
 Plate: 00190 row: d column: 9.
 Location/Qualifiers

FEATURES

source
 1..518
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="nrdg1-00190-d9"
 /cissue_type="Dorsal Root Ganglia"
 /clone_lib="nrdg1 (10855)"
 /note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; rat
 dorsal root ganglia"
 139 c 198 g 78 t

BASE COUNT 103 a 139 c 198 g 78 t

ORIGIN

Query Match 95.0%; Score 19; DB 14; Length 518;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGACACAG 19
 |||||
 DB 407 AGGATGCTGACACAG 425

RESULT 5
 BB084267 251 bp mRNA linear EST 28-JUN-2000
 LOCUS BB084267 RIKEN full-length enriched, adult male dlencephalon Mus
 DEFINITION musculus cDNA clone 930187p21 3', mRNA sequence.
 ACCESSION BB084267
 VERSION BB084267.1 GI:8649585
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 251)

REFERENCE
 1 (bases 1 to 251)
 AUTHORS Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
 P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Igarashi,
 Hirose, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
 Iwawata, T., Kadoya, K., Kagawa, I., Kai, C., Kawai, U., Kikuchi, N.,
 Kiyosawa, H., Kojima, Y., Kondo, S., Koye, S., Kurihara, C., Kusakabe, M.,
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,
 Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sobabe, Y., Sugahara, Y.,
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toyota,
 T., Tsunoda, Y., Watanishi, A., Watanabe, S., Yamamura, T., Yamazaki, I.,
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,
 M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished
 Contact: Yoshinori Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Sueno-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 url: http://genome.gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagakawa, S., Sasaki,
 N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermostabilization and thermoactivation of thermostable enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Iizawa, M., Kawai, J.,
 Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
 Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.
 Location/Qualifiers

FEATURES

Source

```

source
1..251
! .organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="93j0187P21"
/sex="male"
/tissue_type="diencephalon"
/dev_stage="adult"
/lab_host="DH10B"
/clone_1ib="Riken full-length enriched, adult male diencephalon"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGACGAGATCCACAGCTTTTTCCTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGCGATTCCGAGTAAATAAATTAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified plasmid pUC(+ after bulk excision from Lambda phage T. Cloning sites, 5' end: SalI; 3' end: BamHI."

```

BASE COUNT	60 a	52 c	46 g	93 t
ORIGIN				
Query Match		85.0%	Score 17;	DB 10; Length 251;
Best Local Similarity		100.0%	Pred. No. 23;	
Matches 17; Conservative		0;	Mismatches	0; Indels 0; Gaps 0;

RESULT 6	BY649980	400 bp	linear	EST 16-DEC-2002
LOCUS	BY649980			
DEFINITION	BY649980 RIKEN full-length enriched, visual cortex			Mus musculus
ACCESSION	CDNA clone K530046A05.3,			mRNA sequence.
VERSION	BY649980			
KEYWORDS	BY649980.1			GI:27006575
SOURCE	EST.			
ORGANISM	Mus musculus (house mouse)			
	Mus musculus			

REFERENCE
AUTHORS

1 (bases 1 to 400)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Oato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schombach, C., Gotohori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Knapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusic, V., Choitha, C., Cobarr, L. E., Collins, S., Dalla, E., Drazani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Guslichich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Keddeski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lehman, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavani, M. J., Petta, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Remachandran, S., Ravasi, T., Reed, J. C., Reed, D. V., Reid, J., Ring, B. Z., Ringwald, M., Sadelain, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Watanabe, C., Wang, Y.,

Watanabe, Y., Wells, C., Wilmng, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hizcane-Tsitsikawa, T., Komno, H., Nakamura, M., Sakakume, N., Sato, K., Shiraki, T., Waki, K., Kawaji, J., Aizawa, K., Aizawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Saeki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Watanabe, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL
NATURE 420, 563-573 (2002)

MEDLINE
22354663

PUBMED
12466851

COMMENT
Contact: Yoshihide Hayashizaki, yoshihide@nagasaki-nu.ac.jp

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences *Mamm. Genome* 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes *Genome Res.* 10 (10), 1617-1630 (2000)

Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer *Genome Res.* 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library *Genome Res.* 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Michela Pajolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa, Wakoh-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

```

BASE COUNT      119 a      59 c      83 g      139 t
ORIGIN
/organism="Mus musculus"
/mol_type="RNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K530046A05"
/tissue_type="visual cortex"
/clone_idb="RIKEN full-length enriched, visual cortex"

```

Query Match	85.0%;	Score 17;	DB 13;	Length 400;
Best Local Similarity	100.0%;	Pred. No. 27;		
Matches	17;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;
QY	2	GGGATGCTGGACACAA	18	
Db	166	GGGATGCTGGACACAA	182	

RESULT 7
BY640733

LOCUS	DEFINITION
B640733	433 bp mRNA linear EST 15-DEC-2000
B640733	Riken full-length enriched, visual cortex Mus musculus
CYDA0723	cDNA clone K43033806 3', mRNA sequence.
B640733	
B640733.1	GI:26975915
EST.	
Mus musculus (house mouse)	
Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Oatso, N., Satou, R., Suzuki, H., Yamamaki, I., Kiyosawa, H., Nagai, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schombach, C., Gotoori, T., Baldarelli, R., Hill, D.P., Bulic, C., Hune, D.A., Quackenbush, J., Schriml, L.M., Kapran, A., Matsuda, H., Batilov, S., Beisel, R.M., Blake, J.A., Brad, D., Brusic, V., Chothia, C., Corbett, L.E., Cousins, S., Dalla, E., Drean, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gisli, C., Godzik, A., Gough, J., Gimond, S., Gustinchik, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lemnar, B., Lyons, P.A., Maglott, D.R., Maltsev, U., Marchionni, L., McKenzie, J., Mikki, H., Nagashima, T., Numata, K., Okido, T., Pavani, W.J., Petosa, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.W., Q.D., Remachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shmida, K., Sultana, R., Takemura, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Veerardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshak-Boris, A., Yanagisawa, M., Yang, T., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirotsugu-Kishikawa, T., Komno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Akawa, K., Arakawa, T., Fukuda, S., Harai, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	22354683
PUBMED	12466851
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenocho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirotsugu-Kishikawa, T., Kawai, J., Kondo, H., Miyazaki, A., Ohno, M., Sakai, K., Murata, M., Nakamura, M., Nomura, K., Numasaki, R., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Komno, H., Miyazaki, A., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tsgami, M., Waki, K., Watanishi, A., Yamawaku, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10). 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 261-269 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

issues were provided by Mitsuha Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development, Brain Science Institute, RIKEN 2-1 Hiroseawa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES		source		Location/Qualifiers	
		1.	.433		
		/organism="Mus musculus"			
		/mol_type="mRNA"			
		/strain="C57BL/6J"			
		/db_xref="taxon:10090"			
		/clone="K43038006"			
		/tissue_type="visual cortex"			
		/clone_idb="RIKEN full-length enriched, visual cortex"			
BASE COUNT	127	a	66	c	89
ORIGIN				g	150
				t	1
				others	
Query Match		85.0%	Score 17;	DB 13;	Length 433;
Best Local Similarity		100.0%;	Pred No. 27;		
Matches 17;	Conservative	0;	Mismatches	0;	Indels
				Gaps	0;
OY	2	GGGATGCTGGACACAA	18		
DB	198	GGGATGCTGGACACAA	214		

```

RESULT 8
BP931900
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM
SOURCE
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
1. 254 bp mRNA linear EST 22-JAN-2001
BP931900
CM4-NT0213-151200-607-b08 NT0213 Homo sapiens cDNA, mRNA sequence.
BP931900
BP931900.1 GI:123495224
EST.
Homo sapiens (human)
Homo sapiens
Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 254)
Dias Neto, E., Garcia Correa, R., Veitovskii-Almeida, S., Britones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jorgensen, C. V., O'Hare
, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. V. and
Simpson, A. J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01503-010, Sao Paulo-SF,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/ICR Human Cancer Genome
Project. This entry can be seen in the following URL
http://www.ludwig.org.br/scripts/gethtml2.pl?l1=CM4&t2=CM4-NT0213
151200-607-b08&t3=2000-12-15&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 20
High quality sequence stop: 252.
Location/Qualifiers
1. 254
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/def_stage="Adult"
/clone_id="NT0213"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: Sma1;

```

Site 2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 61 a 64 c 54 g 75 t
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 3 GGATGCTGGACACAA 18
95 GGATGCTGGACACAA 110

RESULT 9
A0552781
LOCUS

DEFINITION RPCI-11-414E11, TV RPCI-11 Homo sapiens genomic clone RPCI-11-414E11, genomic survey sequence.

ACCESSION A0552781
KEYWORDS A0552781.1 GI:4911958

VERSION GSS.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 259)

AUTHORS Zhao, S., Adams, M.D., Niernan, W., Malek, J., de Jong, P. and Venter, J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

JOURNAL Map Building

COMMENT Unpublished

Contact: Shaying Zhao, William Niernan, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0208
Fax: 301 838 0208

Email: hbe@igir.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieterdejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet. cs (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.

Seq primer: T7

Class: BAC ends.

FEATURES

source

1..259

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="GDB:7658698"

/cd_xref="taxon:9606"

/clone="RPCI-11-414E11"

/sex="Male"

/cell_type="Lymphocytes"

/clone_id="RPCI-11"

/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;

RPCI11 Human Male BAC library"

BASE COUNT 48 a 89 c 67 g 55 t

ORIGIN

Query Match 80.0%; Score 16; DB 28; Length 259;

Best Local Similarity 100.0%; Pred. No. 84;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 2 GGATGCTGGACACAA 17

138 GGATGCTGGACACAA 153

RESULT 10
AV733084 361 bp mRNA linear EST 17-OCT-2000

LOCUS AV733084 cda Homo sapiens cDNA clone cdaRD10 5', mRNA sequence.

DEFINITION AV733084

VERSION AV733084.1 GI:10850629

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 361)

AUTHORS Yang, Y., Song, H., Peng, Y., Gu, Y., Gao, G., Xiao, H., Xu, X., Li, N.,

Qian, Z., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu

W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J.,

Chen, Z. and Han, Z.

Homo sapiens cDNA cda clones

Unpublished

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

source

1..361

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="cdaRD10"

/tissue_type="pheochromocytoma"

/dev_stage="Adult"

/lab_host="EM25.8"

/clone_id="cda"

/note="Vector: pTRIPLEX2; Site_1: sfIIA; Site_2: sfIIB"

BASE COUNT 100 a 73 c 90 g 98 t

ORIGIN

Query Match 80.0%; Score 16; DB 9; Length 361;

Best Local Similarity 100.0%; Pred. No. 92;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 4 GATGCTGGACACAA 19

114 GATGCTGGACACAA 129

RESULT 11

CB774635 392 bp mRNA linear EST 16-MAY-2003

LOCUS AMGNMNC:NRPI4-00029-E2-A W Rat pituitary (10472) Rattus norvegicus

DEFINITION CB774635

VERSION CB774635.1 GI:29863026

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 392)

AUTHORS Amgen EST Program.

TITLE Amgen Rat EST Program

JOURNAL Unpublished

COMMENT Contact: Dan Fitzpatrick

Amgen, Inc

One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA

Tel: 805 447-4861

Plate: 00029 row: e column: 2.

FEATURES

Location/Qualifiers
1..392
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone_lib="W Rat pituitary (10472)"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; W Rat pituitary adult female Mistar rat avg insert size 2.1 kb"

BASE COUNT 99 a 103 c 105 g 83 t 2 others

Query Match 80.0%; Score 16; DB 14; Length 392;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGATGCTGGACACA 17
DB 40 GGGATGCTGGACACA 55

RESULT 12
LOCUS CB436177/c 638 bp mRNA linear EST 25-MAR-2003
DEFINITION CB436177 616502 MARC 6BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION CB436177
VERSION CB436177.1 GI:29217946
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
1 (bases 1 to 638)
Smith,T.P.L., Roberts,A.J., Echtenkamp,S.E., Chitko-McKown,C.G., Wray,J.E. and Keele,J.W.
A second set of bovine ESTs from pooled-tissue normalized libraries
Unpublished
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail@marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with
Cross_match v0.990329.
Plate: FQY8030 row: L column: 17
Seq primer: GPATACGACCTCCTAATAGG.
Location/Qualifiers
1..638

FEATURES
source
1..638
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 6BOV"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI; library made with RNA pooled from multiple tissues including liver, lung, hypothalamus, pituitary, and placenta/endometrium."

BASE COUNT 185 a 104 c 146 g 203 t

Query Match 80.0%; Score 16; DB 14; Length 638;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GATGCTGGACACAAG 19
DB 27 GATGCTGGACACAAG 12

RESULT 13
LOCUS AG148078 668 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-010C19.TU, genomic survey sequence.
ACCESSION AG148078
VERSION AG148078.1 GI:16677756
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.

REFERENCE
1
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of library RP43-43
Unpublished
2 (bases 1 to 668)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimps@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library RP43-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS

Sequencing: TV
LIBRARY
Vector : PBAC3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
Location/Qualifiers
1..668
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-010C19.TU"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP43-43 Chimpanzee Male BAC Library"

COMMENT

FEATURES

source
1..668
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-010C19.TU"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP43-43 Chimpanzee Male BAC Library"

BASE COUNT 155 a 186 c 175 g 152 t

Query Match 80.0%; Score 16; DB 29; Length 668;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGATGCTGGACACA 17
DB 168 GGGATGCTGGACACA 183

RESULT 14
LOCUS BE544293 785 bp mRNA linear EST 09-AUG-2000
DEFINITION 601076768F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3462593 5', mRNA sequence.
ACCESSION BE544293
VERSION BE544293.1 GI:9772938
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 785)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.

REFERENCE
1
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.

QY 5 ATGCTTGACACAG 19
 DB 68 ATGCTTGACACAG 54
 RESULT 17
 BB281631 304 bp mRNA linear EST 01-AUG-2000
 DEFINITION BB281631 RIKEN full-length enriched, adult retina Mus musculus cDNA
 clone A930032K17 3', mRNA sequence.
 BB281631
 ACCESSION BB281631.1 GI:8982080
 VERSION
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 304)
 REFERENCE
 AUTHORS Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Oho, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomihaga, N., Toya, T., Tsunoda, Y., Watanabe, A., Watanabe, S., Yamamura, T., Yamataka, T., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Kono, H., et al.)
 Unpublished
 TITLE Contact: Yoshitake Hayashizaki
 JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic
 COMMENT Science Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 The most stable and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kishimoto, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.
 Location/Qualifiers
 1..304
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="A930032K17"
 /tissue_type="retina"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, adult retina"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was

BASE COUNT 78 a 51 c 58 g 116 t 1 others
 ORIGIN
 Query Match 75.0%; Score 15; DB 10; Length 304;
 Best Local Similarity 100.0%; Pred. No. 3, 1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGGATGCTTGACAC 16
 DB 76 GGGATGCTTGACAC 90
 RESULT 18
 BQ328868 353 bp mRNA linear EST 17-MAY-2002
 LOCUS R00-EN0018-050600-021-e07 EN0018 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BQ328868
 ACCESSION BQ328868.1 GI:20968839
 VERSION
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 353)
 REFERENCE
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brines, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.U.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC0&t=RC0-EN0018-
 050600-021-e07&t3=2000-06-05&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 32
 High quality sequence stop: 353.
 Location/Qualifiers
 1..353
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="adult"
 /clone_lib="EN0018"
 /note="Organ: lung normal; Vector: puc18; Site 1: SmaI;
 Site 2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 19 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

Query Match 75.0%; Score 15; DB 13; Length 353;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGACA 15
|||||
148 AGGATGCTGACA 162

RESULT 19
BOJ28810/c 366 bp mRNA linear EST 17-MAY-2002
LOCUS RCO-EN0024-120500-031-h02 EN0024 Homo sapiens cDNA, mRNA sequence.
DEFINITION BOJ28810
ACCESSION BOJ28810.1 GI:20968740
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 366)
Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,M.Jr., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brennan,R.K., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
JOURNAL
MEDLINE
PubMed
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RCO&t2=RCO-EN0024-120500-031-h02&t3=2000-05-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 4
High quality sequence stop: 366.
Location/Qualifiers
1..366
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/def_star="Adult"
/clone_lib="EN0024"
/note="Organ: lung_normal; Vector: puc18; Site: 1; Smal; Site 2: Smal; A mini library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 44 a 146 c 90 g 86 t
ORIGIN
Query Match 75.0%; Score 15; DB 13; Length 366;

Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGACA 15
|||||
302 AGGATGCTGACA 288

RESULT 20
BY608974 368 bp mRNA linear EST 15-DEC-2002
LOCUS BY608974 RIKEN full-length enriched, visual cortex Mus musculus
DEFINITION BY608974 RIKEN full-length enriched, visual cortex Mus musculus
ACCESSION BY608974
VERSION BY608974.1 GI:26944156
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 368)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Boro,H., Kondo,S., Nikaide,I., Osato,N., Salto,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Butt,C., Hume,D.A., Quackenbush,J., Schmitt,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chochia,C., Corbett,L.B., Cousins,S., Dalia,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Fraser,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godik,A., Gough,V., Grimond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawai,H., Kawasawa,Y., Kedzierski,R.W., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lennard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pereira,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramchandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sanderlin,A., Schneider,C., Sample,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wyshaw-Boris,A., Yangisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Koshikawa,T., Kono,H., Nakamura,M., Sakakume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Aizawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE
JOURNAL
MEDLINE
PubMed
COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,Aizawa,K., Akimura,T., Arakawa,T., Kawai,D., Kono,H., Miyazaki,A., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Michela Fagiolini and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0196 Japan) whose assistance we gratefully acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES
 source
 1. 368
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="X23030CN15"
 /tissue_type="visual cortex"
 /clone_lib="RIKEN full-length enriched, visual cortex"
 /clone_1b="89 a 84 c 105 g 89 t 1 others"

BASE COUNT
 89 a 84 c 105 g 89 t 1 others
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGGATGCTGGACA 15
 |||||
 Db 267 AGGGATGCTGGACA 281

RESULT 21
 BQ490402/c 420 bp mRNA linear EST 07-JUN-2002
 LOCUS 07-B011789-006-050-N13-T3 Sugar beet MP1Z-ADIS-006 Lambda Zap II
 DEFINITION 118189 Beta vulgaris cDNA clone N-13-50, mRNA sequence.
 ACCESSION BQ490402
 VERSION BQ490402.1 GI:21335022
 KEYWORDS EST.
 SOURCE Beta vulgaris
 ORGANISM Beta vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Amaranthaceae; Beta.

REFERENCE
 AUTHORS Bellin, D., Werber, M., Theis, T., Weishaar, B. and Schneider, K.
 TITLE EST sequencing, annotation and macroarray expression analysis of more than 3000 sugar beet cDNAs identifies genes with root-specific expression pattern

JOURNAL
 COMMENT Unpublished
 CONTACT: Weishaar B
 ADIS DNA core facility at MP1Z
 Max-Planck-Institute for Plant Breeding Research
 Carl-von-Linne-Weg 10, 50823 Koeln, Germany
 Fax: 00492215062851
 Email: weishaar@mp1z-koeln.mpg.de
 Seq primer: 13 /ATTTACCTCTCACTAAGG/
 High quality sequence stop: 420.
 Location/Qualifiers
 1. 420

FEATURES
 source
 /organism="Beta vulgaris"
 /mol_type="mRNA"
 /db_xref="taxon:161934"
 /clone="N-13-50"
 /dev_stage="4 week old pot-grown plants"
 /clone_1b="Sugar beet MP1Z-ADIS-006 Lambda Zap II library"

/note="Organ: shoot and root; Vector: pBluescript SK- from lambda Zap II; cDNA (lambda Zap-II) library from sugar beet, whole plant mRNA. Prepared using the Stratagene UniZAP cDNA kit; cloning sites EcoRI-XhoI, primer sites and orientation: rev-T3-SacI-SK-EcoRI-GGACGAGG-5pr-cDNA-polyA-XhoI-KpnI-T7-unl"
 BASE COUNT
 102 a 71 c 98 g 149 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Pred. No. 3.4e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TGCCTGGACACAGA 20
 |||||
 Db 183 TGCCTGGACACAGA 169

RESULT 22
 B49246/c 440 bp DNA linear GSS 08-APR-1999
 LOCUS R0C111-5J13.TP RPCI-11 Homo sapiens genomic clone RPCI-11-5J13,
 DEFINITION genomic survey sequence.
 ACCESSION B49246
 VERSION B49246.1 GI:2601483
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 440)
 REFERENCE
 AUTHORS Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden
 'K., Berry, K., Grainger, D., Suh, E., Wible, C., de Jong, P. and Venter
 , U.C.
 TITLE Use of BAC End Sequences for Sequence-Ready Map Building

JOURNAL
 COMMENT Unpublished
 CONTACT: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@igf.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://info@resgen.com>). BAC end search page: http://www.igf.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers
 1. 440

FEATURES
 source
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="GDB:7501764"
 /db_xref="taxon:9606"
 /clone="RPCI-11-5J13"
 /sex="Male"
 /cell_type="lymphocytes"
 /clone_1b="RPCI-11"
 /note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;
 RPCI11 Human Male BAC library"
 RPCI11 Human Male BAC library"

BASE COUNT
 72 a 128 c 122 g 118 t
 ORIGIN

Query Match
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ATGCTGGACACAGA 19
 |||||

Mon Aug 18 10:30:10 2003

us-10-074-620-5.cll.rst

Page 12

Db 431 ATGCTGACACAGA 417

RESULT 23
AL641014
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL641014 446 bp mRNA linear EST 12-DEC-2001
AL641014 XGC-neurula Silurana tropicalis cDNA clone TNeu026109.5',
mRNA sequence.
AL641014 GI:16793139
EST.
Silurana tropicalis (western clawed frog)
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodidae; Silurana.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Huckie E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu026109.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..446
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TNeu026109"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5' of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into PCS107 with
EcoRI at the 5' end and NotI at the 3' end."

FEATURES
source
1..446
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TNeu026109"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: PCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from 5' of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into PCS107 with
EcoRI at the 5' end and NotI at the 3' end."

BASE COUNT 159 a 86 c 71 g 130 t

Query Match 75.0%; Score 15; DB 9; Length 446;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 TGCTGACACAGA 20
Db 251 TGCTGACACAGA 265

RESULT 24
CB787356 451 bp mRNA linear EST 16-MAY-2003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CB787356 451 bp mRNA linear EST 16-MAY-2003
AMGNNTC:SRPB2-00216-C2-A srpb2 (10220) Rattus norvegicus cDNA clone
srpb2-00216-C2.5', mRNA sequence.
CB787356
CB787356.1 GI:29875747
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 451)
Amgen EST Program.
Amgen Rat EST Program
Unpublished
Contact: Dan Fitzpatrick
Amgen, Inc

One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00216 row: c column: 2.
Location/Qualifiers
1..451
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="srpb2-00216-C2"
/issue_type="prostate tissue"
/clone_id="srpb2 (10220)"
/note="Vector: PSPORI; Site 1: SalI; Site 2: NotI; rat
prostate normalized double selected poly(A+) mRNA size
fraction > 1 kb"

BASE COUNT 122 a 103 c 87 g 134 t 5 others

Query Match 75.0%; Score 15; DB 14; Length 451;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 TGCTGACACAGA 20
Db 66 TGCTGACACAGA 52

RESULT 25
A0359962 520 bp DNA linear GSS 06-MAR-1999
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

A0359962 520 bp DNA linear GSS 06-MAR-1999
HS_5034_B1_c11_T7 RPII1 Human Male BAC Library Homo sapiens
genomic clone Plate=610 Col=21 Row=F, genomic survey sequence.
A0359962
A0359962.1 GI:4208838
GSS.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 520)
Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams W.D. and
Hood L.

Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
10449764
Contact: Mahairas G.G., Wallace J.C., Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 610 row: F column: 21
Seq primer: 17
Classes: BAC ends
High quality sequence stop: 520.
Location/Qualifiers
1..520
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=610 Col=21 Row=F"
/sex="Male"
/cell_type="Lymphocytes"
/clone_id="RPII1 Human Male BAC Library"
/note="Vector: pBAC3.6; RPII1 Human Male BAC Library"

BASE COUNT 128 a 124 c 121 g 140 t 7 others


```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE598D0911446 ; IMAGE:5178968"
/lab_host="DH10B"
/clone_lib="NIH_MGC_115"
/note="Organ: pooled brain, lung, testis; Vector:
pcwv-sport6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dt primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."

BASE COUNT      127 a      147 c      185 g      76 t
ORIGIN

Query Match      75.0%; Score 15; DB 13; Length 535;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGGATGCTCGACACA 15
|||||
256 AGGATGCTCGACACA 270

RESULT 29
BB641771      561 bp      mRNA      linear      EST 31-AUG-2001
LOCUS      BB641771      RIKEN full-length enriched, 10 days neonate cortex Mus
DEFINITION      musculus cDNA clone AB3004815 5', mRNA sequence.
ACCESSION      BB641771
VERSION      BB641771.1 GI:15401716
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
1 (bases 1 to 561)
Arakawa,T., Carrincci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hizemoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Komoto,H., Kouda
,M., Koyama,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Sasaki
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y. et al. 2001)
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished
Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carrincci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Komoto,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Komoto,H., Fukunishi,Y., Shibata,K., Itoh,M., Carrincci,P., Sugahara
,Y. and Hayashizaki,Y.

```

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T., Ishii,Y. and Hayashizaki,Y.

Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, U72-L86 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

```

FEATURES
source
location/Qualifiers
1..561
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="AB3004815"
/tissue_type="cortex"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 10 days neonate
cortex"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 15'
GAGGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTT 3', cDNA was
prepared by using trihalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence 15' GAGGAGAGATTCGATTAATTAATTAATTCCTCCCCCCCC
3'. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pluscript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT      121 a      160 c      156 g      124 t
ORIGIN

Query Match      75.0%; Score 15; DB 10; Length 561;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      5 ATGCTGACACACAG 19
|||||
264 ATGCTGACACACAG 278

RESULT 30
BQ489926/c      603 bp      mRNA      linear      EST 07-JUN-2002
LOCUS      BQ489926      20-59428-006-009-J20-T3 Sugar beet MP12-ADIS-006 Lambda Zap II
DEFINITION      library Beta vulgaris cDNA clone J-20-9, mRNA sequence.
ACCESSION      BQ489926
VERSION      BQ489926.1 GI:21334546
KEYWORDS      EST.
SOURCE      Beta vulgaris
ORGANISM      Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllales; Amaranthaceae; Beta.
1 (bases 1 to 603)
Bellin,D., Weber,M., Theis,T., Weishaar,B. and Schneider,K.
EST sequencing, annotation and macroarray expression analysis of
more than 3000 sugar beet cDNAs identifies genes with root-specific
expression pattern
Unpublished
Contact: Weishaar B
ADIS DNA core facility at MPIZ
Max-Planck-Institute for Plant Breeding Research

```

Carl-von-Linne Weg 10, 50829 Koeln, Germany
 Fax: 00492215062851
 Email: weisshaampitz-koeln.mpg.de
 Seq primer: T3 'AATTAACCTCCTCAACAGG'
 High quality sequence stop: 603.
 Location/Qualifiers

FEATURES

source

1. 603
 /organism="Beta vulgaris"
 /mol_type="mRNA"
 /db_xref="taxon:161934"
 /clone="J-20-9"
 /dev_stage="4 week old pot-grown plants"
 /clone_lib="Sugar Beet MP1Z-ADIS-006 Lambda Zap II library"
 /note="Organ: shoot and root; Vector: pBluescript SK- from lambda Zap II; CDNA (lambda Zap-II) library from sugar beet, whole plant mRNA, prepared using the Stratagene UnizAP CDNA kit, cloning sites EcoRI-XhoI, primer sites and orientation:
 rev-T3-Sect-SK-BcoRI-GGCGACGAG-5pr-cDNA-polyA-XhoI-KpnI-T7-unl"

BASE COUNT 154 a 126 c 133 g 190 t
 ORIGIN

Query Match 75.0%; Score 15; DB 13; Length 603;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 TGCTTGACACAGA 20
 |||||
 Db 219 TGCTTGACACAGA 205

RESULT 31
 B0591046/c 603 bp mRNA linear EST 06-DEC-2002
 LOCUS E012597-024-018-G02-SP6 MP1Z-ADIS-024-storage root Beta vulgaris
 DEFINITION CDNA clone 024-018-G02 5-PRIME, mRNA sequence.
 ACCESSION B0591046
 VERSION B0591046.1 GI:26120629
 KEYWORDS EST.
 SOURCE Beta vulgaris
 ORGANISM Beta vulgaris

REFERENCE
 AUTHORS Herwig,R., Schulz,B., Weisshaar,B., Hennig,S., Steinfath,M., Drungowski,M., Stahl,D., Wruick,W., Menze,A., O'Brien,J., Lehnach,H. and Radelof,U.
 TITLE Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes
 JOURNAL Plant J. 32 (5), 845-857 (2002)
 COMMENT Contact: Weisshaar B
 ADIS DNA core facility at MP1Z
 Max-Planck-Institute for Plant Breeding Research
 Carl-von-Linne Weg 10, 50829 Koeln, Germany
 Fax: 00492215062851
 Email: weisshaampitz-koeln.mpg.de
 Insert Length: 603 Std Error: 0.00
 Plate: 18 row: G column: 02
 Seq primer: SP6: CATACGATTGCGTACCTATAG.
 Location/Qualifiers

1. 603
 /organism="Beta vulgaris"
 /mol_type="mRNA"
 /cultivar="KWS2320 (double haploid, monogerm breeding line)"
 /db_xref="GABI:189076"
 /db_xref="taxon:161934"
 /clone="024-018-G02"
 /tissue_type="storage root"
 /lab_host="EMDHI0B"

FEATURES
 source

1. 603
 /organism="Beta vulgaris"
 /mol_type="mRNA"
 /cultivar="KWS2320 (double haploid, monogerm breeding line)"
 /db_xref="GABI:189076"
 /db_xref="taxon:161934"
 /clone="024-018-G02"
 /tissue_type="storage root"
 /lab_host="EMDHI0B"

BASE COUNT 163 a 121 c 128 g 191 t
 ORIGIN

Query Match 75.0%; Score 15; DB 13; Length 603;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 TGCTTGACACAGA 20
 |||||
 Db 198 TGCTTGACACAGA 184

RESULT 32
 A0157654/c 624 bp DNA linear GSS 12-SEP-1998
 LOCUS nbxb000906r CUGI Rice BAC Library Oryza sativa (japonica
 DEFINITION cultivar-group) genomic clone nbxb000906r, genomic survey
 sequence.
 ACCESSION A0157654
 VERSION A0157654.1 GI:3554679
 KEYWORDS GSS.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)

REFERENCE
 AUTHORS Wing,R.A. and Dean,R.A.
 TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
 JOURNAL Unpublished
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seg primer: GGAACAGCTATGACCATG
 Class: BAC ends
 High quality sequence start: 45
 High quality sequence stop: 365.
 Location/Qualifiers

1. 624
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /strain="japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="nbxb000906r"
 /tissue_type="leaf"
 /lab_host="E. coli DH10B"
 /clone_lib="CUGI Rice BAC Library"
 /note="Vector: pBelBAC11; Site 1: HindIII; Site 2:
 HindIII; Rice is one of two most popular grains in the
 world. Half of the world population especially those
 inhabiting highly populated areas of the humid tropics
 and subtropics, rely on rice as their primary source of
 carbohydrate. Monocotyledonous rice is a diploid plant
 (2n=24) with a haploid genome equivalent of 431 Mbp
 (Arumuganathan and Earle, 1991). The relatively small
 genome of rice, three times larger than that of
 Arabidopsis, makes it suitable for genomic studies. In

order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 171 a 153 c 121 g 179 t
 ORIGIN
 Query Match 75.0%; Score 15; DB 28; Length 624;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 5 ATGCTGACACAG 19
 Db 432 ATGCTGACACAG 418

RESULT 33
 BM725133 629 bp mRNA linear EST 01-MAR-2002
 LOCUS UI-E-B01-aj-a-p-24-0-UI_r1 UI-E-B01 Homo sapiens cDNA clone
 DEFINITION UI-E-B01-aj-a-p-24-0-UI 5', mRNA sequence.
 ACCESSION BM725133
 VERSION BM725133.1 GI:19046464
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia, Eutheria; Primates; Carnivora; Hominiidae; Homo.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548

COMMENT
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@iowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
 Seq primer: M3 Reverse.
 Location/Qualifiers
 1..629
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-B01-aj-a-p-24-0-UI"
 /tissue_type="fetal eye"
 /dev_stage="fetal"
 /lab_host="MD10B (Life Technologies) (T1 phase resistant)"
 /clone_lib="UI-E-B01"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-B01 is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally

into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCGGATACCC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."
 BASE COUNT 120 a 181 c 208 g 120 t
 ORIGIN
 Query Match 75.0%; Score 15; DB 12; Length 629;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 3 GGATGCTGACACA 17
 Db 162 GGATGCTGACACA 148

RESULT 34
 CB918954 632 bp mRNA linear EST 25-APR-2003
 LOCUS VVD039E12.348541 An expressed sequence tag database for abiotic stressed berries of Vitis vinifera var. Chardonnay
 DEFINITION CDNA clone VVD039E12 5', mRNA sequence.
 ACCESSION CB918954
 VERSION CB918954.1 GI:30133615
 KEYWORDS EST.
 SOURCE Vitis vinifera
 ORGANISM Vitis vinifera

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 TITLE Vitis vinifera var. Chardonnay
 JOURNAL Unpublished
 COMMENT Contact: Cushman JC
 Department of Biochemistry
 University of Nevada
 MS200, Reno, NV 89557-0014, USA
 Tel: 775-784-1918
 Fax: 775-784-1650
 Email: jcushman@unr.edu
 PCR Primers
 FORWARD: T3 20mer
 BACKWARD: T7 21mer (backward)
 Plate: 039 row: E column: 12
 Seq primer: T3 20mer
 High quality sequence stop: 632.
 Location/Qualifiers
 1..632
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /db_xref="taxon:29760"
 /clone="VVD039E12"
 /tissue_type="berries"
 /dev_stage="mixed 8, 9, 11, 13, 15, 16 weeks dat"
 /clone_lib="An expressed sequence tag database for abiotic stressed berries of Vitis vinifera var. Chardonnay"
 /note="Vector: lambda Uni-Zap XR, Bluescript SK-; Site_1: EcoRI; Site_2: XhoI"

FEATURES
 source
 1..632
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /db_xref="taxon:29760"
 /clone="VVD039E12"
 /tissue_type="berries"
 /dev_stage="mixed 8, 9, 11, 13, 15, 16 weeks dat"
 /clone_lib="An expressed sequence tag database for abiotic stressed berries of Vitis vinifera var. Chardonnay"
 /note="Vector: lambda Uni-Zap XR, Bluescript SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 179 a 128 c 159 g 166 t
 ORIGIN

Query Match 75.0%; Score 15; DB 14; Length 632;
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 GGATGCTGACAC 16
 Db 79 GGATGCTGACAC 93

```

RESULT 35
CB919968
LOCUS
DEFINITION CB919968 642 bp mRNA linear EST 25-APR-2003
stressed berries of Vitis vinifera var. Chardonnay Vitis vinifera
CDNA clone VVD057F10 5, mRNA sequence.
ACCESSION CB919968
VERSION CB919968.1 GI:30134630
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; Vitaceae; Vitis.
1 (bases 1 to 642)
Cushman,J.C.
An expressed sequence tag database for abiotic stressed berries of
Vitis vinifera var. Chardonnay
Unpublished
JOURNAL
Contact: Cushman JC
Department of Biochemistry
University of Nevada
MS200, Reno, NV 89557-0014, USA
Tel: 775-784-1918
Fax: 775-784-1650
Email: jcushman@unr.edu
PCR Primers
FORWARD: T3 20mer
BACKWARD: T7 21mer (backward)
Plate: 057 row: F column: 10
Seq primer: T3 20mer
High quality sequence scorp: 642.
location/Qualifiers
1..642
source
/organism="Vitis vinifera"
/mol_type="mRNA"
/db_xref="taxon:29760"
/clone="VVD057F10"
/tissue_type="berries"
/dev_stage="mixed: 8, 9, 11, 13, 15, 16 weeks daf"
/clone_lib="An expressed sequence tag database for abiotic
stressed berries of Vitis vinifera var. Chardonnay"
/notes="Vector: Lambda Uni-zap XR, Bluescript SK-; Site_1:
EcorI; Site_2: XhoI"
BASE COUNT 180 a 130 c 161 g 168 t 3 others
ORIGIN
Query Match 75.0%; Score 15; DB 14; Length 642;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GGGATGCTGTGACAC 16
Db 79 GGGATGCTGTGACAC 93

```

Search completed: August 15, 2003, 10:57:57
 Job time : 1261 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:32:41 ; Search time 547.75 Seconds
(without alignments)
1493.734 Million cell updates/sec

Title: US-10-074-620-5

Perfect score: 20

Sequence: 1 agggatgcctggacacacaga 20

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 2888711 segs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_un:*
28: em_vt:*
29: em_vt:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_mus:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rnd:*
36: em_hcg_rnd:*
37: em_hcg_vit:*
38: em_sy:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_oher:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AX522240	AX522240 Sequence
2	20	100.0	5050	HS4021R2A	K03333 Epstein-Bar
3	20	100.0	5311	HS4021R2	K03332 Epstein-Bar
4	20	100.0	171823	HHV507799	AJ507799 Human her
5	20	100.0	172281	EBV	V01555 Epstein-Bar
6	20	100.0	164113	HS4B568RAJ	M08517 Epstein-Bar
7	19	95.0	6076	AF272661	AF272661 Rattus no
8	19	95.0	240584	AC116063	AC116063 Rattus no
9	19	95.0	254452	AC118348	AC118348 Rattus no
10	17	85.0	1612	AF183139	AF183139 Cercopitn
11	17	85.0	235357	AC103758	AC103758 Homo sapi
12	17	85.0	246186	AC095487	AC095487 Rattus no
13	17	85.0	257316	AC095181	AC095181 Rattus no
14	17	85.0	278519	AC119530	AC119530 Rattus no
15	17	85.0	297264	AC099161	AC099161 Rattus no
16	16	80.0	1377	AF542091	AF542091 Oncorhync
17	16	80.0	1452	AX705192	AX705192 Sequence
18	16	80.0	1452	AX705202	AX705202 Sequence
19	16	80.0	1458	AX705204	AX705204 Sequence
20	16	80.0	1545	ES1824	ES1824 Novel amyla
21	16	80.0	1545	ES1825	ES1825 Novel amyla
22	16	80.0	1545	ES1826	ES1826 Novel amyla
23	16	80.0	1545	ES1827	ES1827 Novel amyla
24	16	80.0	1972	A20154	A20154 alpha-amyla
25	16	80.0	2084	BACAM	BACAM Bacillus am
26	16	80.0	2084	AR008285	AR008285 Sequence
27	16	80.0	2084	AR037275	AR037275 Sequence
28	16	80.0	2084	AR052145	AR052145 Sequence
29	16	80.0	2084	AR087551	AR087551 Sequence
30	16	80.0	2084	AR129916	AR129916 Sequence
31	16	80.0	2084	AR137905	AR137905 Sequence
32	16	80.0	2084	AR143262	AR143262 Sequence
33	16	80.0	2084	AR224261	AR224261 Sequence
34	16	80.0	2084	AR285306	AR285306 Sequence
35	16	80.0	2084	AX244195	AX244195 Sequence
36	16	80.0	2084	AX305020	AX305020 Sequence
37	16	80.0	2084	AX332826	AX332826 Sequence
38	16	80.0	2084	AX370717	AX370717 Sequence
39	16	80.0	2084	AX601810	AX601810 Sequence
40	16	80.0	2084	BD077082	BD077082 Alpha-amy
41	16	80.0	2084	BD085805	BD085805 Alpha-amy
42	16	80.0	2604	AR168304	AR168304 Sequence
43	16	80.0	2604	AR215288	AR215288 Sequence
44	16	80.0	2604	AR215288	AR215288 Sequence
45	16	80.0	2604	BD137366	BD137366 Enzymatic
46	16	80.0	60837	HS4354112	HS4354112 Human DNA
47	16	80.0	67962	AL928684	AL928684 Homo sapi
48	16	80.0	73094	AL671114	AL671114 Homo sapi
49	16	80.0	75001	AC018581	AC018581 Homo sapi
50	16	80.0	75001	AC018581	AC018581 Homo sapi
51	16	80.0	93923	AC090818	AC090818 Homo sapi
52	16	80.0	108768	AC105053	AC105053 Homo sapi
53	16	80.0	140542	AC079917	AC079917 Homo sapi
54	16	80.0	147344	AC008907	AC008907 Homo sapi
55	16	80.0	153792	AC099404	AC099404 Oryza sat
56	16	80.0	154244	AL360231	AL360231 Human DNA
57	16	80.0	172683	AC104570	AC104570 Homo sapi
58	16	80.0	174186	AC080027	AC080027 Homo sapi
59	16	80.0	183870	AC115284	AC115284 Homo sapi
60	16	80.0	184158	AC082728	AC082728 Homo sapi
61	16	80.0	186218	AC060831	AC060831 Homo sapi
62	16	80.0	195582	AC105486	AC105486 Rattus no
63	16	80.0	196808	AC009987	AC009987 Homo sapi
64	16	80.0	209252	AC097637	AC097637 Homo sapi
65	16	80.0	215105	AC073717	AC073717 Mus muscu

Pred. No. is the number of results predicted by chance to have a

GTHQCGAGDSPTGSPNSAAPVCNGSHATPNTVPPIHEDESNHSPEADILPDPDWYD
 SITDPADLHESMDYIFFTTSSSDSDVVBSPSRPRPSTQ"

repeat_region 3581..3618
 /note="IR2 repeat partial copy"
 repeat_region 3519..3743
 /note="IR2 repeat copy A"
 repeat_region 3744..3868
 /note="IR2 repeat copy B"
 repeat_region 3869..3993
 /note="IR2 repeat copy C"
 repeat_region 3994..4118
 /note="IR2 repeat copy D"
 repeat_region 4119..4243
 /note="IR2 repeat copy E"
 repeat_region 4244..4368
 /note="IR2 repeat copy F"
 repeat_region 4369..4493
 /note="IR2 repeat copy G"
 repeat_region 4494..4618
 /note="IR2 repeat copy H"
 repeat_region 4619..4743
 /note="IR2 repeat copy I"
 repeat_region 4744..4868
 /note="IR2 repeat copy J"
 repeat_region 4869..4993
 /note="IR2 repeat copy K"
 repeat_region 4994..5118
 /note="IR2 repeat copy L"

BASE COUNT 651 a 1830 c 1417 g 592 t
 ORIGIN 1 bp upstream of BamHI site.

Query Match 100.0%; Score 20; DB 14; Length 5050;
 Best Local Similarity 100.0%; Pred. No. 0.024;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGGATGCTGGACACAGA 20
 Db 1813 AGGGATGCTGGACACAGA 1832

RESULT 3
 HS4U2IR2
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

HS4U2IR2 5311 bp DNA linear VRL 02-AUG-1993
 Epstein-Barr virus (AG876 isolate) U2-IR2 domain encoding nuclear
 protein EBNA2, complete cds.
 K03332 GI:330441
 K03332.1 GI:330441
 nuclear protein.
 Human herpesvirus 4 (Epstein-Barr virus)
 Human herpesvirus 4
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 Gammaherpesvirine; lymphocryptovirus.

1 (bases 1 to 3754: 5256 to 5311)
 Dambaugh, T., Hennessy, R., Chammanak, L. and Kieff, E.
 U2 region of Epstein-Barr virus DNA may encode Epstein-Barr nuclear
 antigen 2
 Proc. Natl. Acad. Sci. U.S.A. 81 (23), 7632-7636 (1984)

JOURNAL MEDLINE
 MEDLINE PUBMED
 85063846
 6209719
 2 (bases 1 to 5311)
 Dambaugh, T.R. and Kieff, E.D.
 Unpublished (1985)
 Original source text: Epstein-Barr virus (AG876 isolate) DNA, clone
 pD113.
 Draft entry and clean copy sequence for [1] kindly provided by
 T.R. Dambaugh, 15-AUG-1985.
 Since the publication of [1] the authors have experimentally
 determined that the gene reported here indeed codes for nuclear
 protein EBNA2 (personal communication, 30-JUN-1986). The base
 polymorphisms found in the IR2 repeats are 80-90% 'a' at position
 3640 and 'c' at 3728 and at the equivalent positions in the other
 IR2 repeats. (Personal communication, 30-JUN-1986). The number of
 IR2 repeats varies between EBV isolates. The strain reported here
 has typically 13 copies. The U2 and U3 unique sequence domains are
 located at positions 634-3591 and 5255-5311. Poly-adenylation

FEATURES	signals
source	location/Qualifiers 1..5311 /organism="Human herpesvirus 4" /mol_type="genomic DNA" /db_xref="taxon:10376"
repeat_region	<1..633 /note="IR1 repeat"
CDS	1512..2876 /note="nucleic acid antigen 2" /codon_start=1 /protein_id="AA45902.1" /db_xref="GI:330442"
repeat_region	/translation="WPYYLAHGGGSYLLVDTMGNGPSISVIFPNPYEQSLNNP LIQIVAGNTGAPAPGPPEPPPEPPPERDWTQPLPLDNWPIGSDASGGPLA SIIMLCMAOYLRLNNAQGGLRPLGPGTRQVTLERQPVNPRQEPAILLOSAPA RFPRFVVNALGHLOCTPREPTLPQPRIPIILPRHNCPATRPAPARLTHGQC LSLPHPHVPHOSTFHCSDBSTGLPPTVSYSISMULSEEPAPPAAAHLPGLVIDYD QQALPRTGPPPMVPVNDPPTTQTPPTKTKGPOGGGRGRGRSGKRGMMKLK DEPRRPDPSPSPMOLPSVNLHOGCGPENSPTBGSTAGTCVRTRPATPLIPSI HEPSSSBSEPPIFLPSDMITPLEPAEDBSMEGIFETTESHSDEENVGPKRRRR TSIQ
repeat_region	3592..3629 /note="IR2 partial copy"
repeat_region	3630..3754 /note="IR2 repeat copy A"
variation	3640 /note="nucleotide polymorphism a or g"
variation	3728 /note="nucleotide polymorphism t or a"
repeat_region	3755..3879 /note="IR2 repeat copy B"
variation	3765 /note="nucleotide polymorphism a or g"
variation	3853 /note="nucleotide polymorphism t or a"
repeat_region	3880..4004 /note="IR2 repeat copy C"
variation	3890 /note="nucleotide polymorphism a or g"
variation	3978 /note="nucleotide polymorphism t or a"
repeat_region	4005..4129 /note="IR2 repeat copy D"
variation	4015 /note="nucleotide polymorphism a or g"
variation	4103 /note="nucleotide polymorphism t or a"
repeat_region	4130..4254 /note="IR2 repeat copy E"
variation	4140 /note="nucleotide polymorphism a or g"
variation	4228 /note="nucleotide polymorphism t or a"
repeat_region	4255..4379 /note="IR2 repeat copy F"
variation	4265 /note="nucleotide polymorphism a or g"
variation	4353 /note="nucleotide polymorphism t or a"
repeat_region	4380..4504 /note="IR2 repeat copy G"
variation	4390 /note="nucleotide polymorphism a or g"
variation	4478 /note="nucleotide polymorphism t or a"
repeat_region	4505..4629 /note="IR2 repeat copy H"
variation	4515 /note="nucleotide polymorphism a or g"
variation	4603 /note="nucleotide polymorphism t or a"
repeat_region	4630..4754

```

variation /note="IR2 repeat copy I"
variation 4640 /note="nucleotide polymorphism a or g"
variation 4778 /note="nucleotide polymorphism t or a"
repeat_region 4755.14879 /note="IR2 repeat copy J"
variation 4765 /note="nucleotide polymorphism a or g"
variation 4853 /note="nucleotide polymorphism t or a"
repeat_region 4880.5004 /note="IR2 repeat copy K"
variation 4890 /note="nucleotide polymorphism a or g"
variation 4978 /note="nucleotide polymorphism t or a"
repeat_region 5005.5129 /note="IR2 repeat copy L"
variation 5015 /note="nucleotide polymorphism a or g"
variation 5103 /note="nucleotide polymorphism t or a"
repeat_region 5130.5254 /note="IR2 repeat copy M"
variation 5140 /note="nucleotide polymorphism a or g"
variation 5228 /note="nucleotide polymorphism t or a"
BASE COUNT 898 a 1864 c 1540 g 1009 t
ORIGIN 1 bp upstream of BamHI site.
Query Match 100.0%; Score 20; DB 14; Length 5311;
Best Local Similarity 100.0%; Pred. NO. 0.024; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

OY 1 AGGATGCTCGACACAGA 20
DB 1740 AGGATGCTCGACACAGA 1759

RESULT 4
LOCUS HHV507799 171823 bp DNA circular VRL 04-APR-2003
DEFINITION Human herpesvirus 4 complete wild type genome.
ACCESSION AJ507799
VERSION AJ507799.1 GI:23893576
KEYWORDS complete genome.
SOURCE Human herpesvirus 4 (Epstein-Barr virus)
ORGANISM Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.

REFERENCE
AUTHORS Arrand, J.R., Rymo, L., Walsh, J.E., Bjorck, E., Lindahl, T. and
Griffin, B.E.
TITLE Molecular cloning of the complete Epstein-Barr virus genome as a
set of overlapping restriction endonuclease fragments
JOURNAL Nucleic Acids Res. 9 (13), 2999-3014 (1981)
MEDLINE 82014887
PUBMED 6269068
REFERENCE
AUTHORS Kozak, M.
TITLE Possible role of flanking nucleotides in recognition of the AUG
initiator codon by eukaryotic ribosomes
JOURNAL Nucleic Acids Res. 9 (20), 5253-5262 (1982)
MEDLINE 82059504
PUBMED 7301588
REFERENCE
AUTHORS Deininger, P.L., Bankier, A., Farrell, P., Baer, R. and Barrell, B.
TITLE Sequence analysis and in vitro transcription of portions of the
Epstein-Barr virus genome
JOURNAL J. Cell. Biochem. 19 (3), 267-274 (1982)
MEDLINE 83109311

```

```

PUBMED 6296170
REFERENCE
AUTHORS Farrell, P.J., Deininger, P.L., Bankier, A. and Barrell, B.
TITLE Homologous upstream sequences near Epstein-Barr virus promoters
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)
MEDLINE 83169725
PUBMED 6300857
REFERENCE
AUTHORS Bankier, A.T., Deininger, P.L., Farrell, P.J. and Barrell, B.G.
TITLE Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8
Epstein-Barr virus
JOURNAL Mol. Biol. Med. 1 (1), 21-45 (1983)
MEDLINE 85035713
PUBMED 6092825
REFERENCE
AUTHORS Seguin, C., Farrell, P.J. and Barrell, B.G.
TITLE DNA sequence and transcription of the BamHI fragment B region of
B95-8 Epstein-Barr virus
JOURNAL Mol. Biol. Med. 1 (3), 369-392 (1983)
MEDLINE 85060424
PUBMED 6094953
REFERENCE
AUTHORS Jeang, K.T. and Hayward, S.D.
TITLE Organization of the Epstein-Barr virus DNA molecule. III. Location
of the P3HR-1 deletion junction and characterization of the NotI
repeat units that form part of the template for an abundant
12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
JOURNAL J. Virol. 48 (1), 135-148 (1983)
MEDLINE 83294686
PUBMED 6310141
REFERENCE
AUTHORS Bankier, A.T., Deininger, P.L., Satchwell, S.C., Baer, R., Farrell, P.J.
and Barrell, B.G.
TITLE DNA sequence analysis of the EcoRI Dhet fragment of B95-8
Epstein-Barr virus containing the terminal repeat sequences
Mol. Biol. Med. 1 (4), 425-445 (1983)
MEDLINE 85060428
PUBMED 6094955
REFERENCE
AUTHORS Farrell, P.J., Bankier, A., Seguin, C., Deininger, P. and Barrell, B.G.
TITLE Latent and lytic cycle promoters of Epstein-Barr virus
JOURNAL EMBO J. 2 (8), 1331-1338 (1983)
MEDLINE 20311131
PUBMED 10872327
REFERENCE
AUTHORS Jones, M.D., Foster, L., Sheedy, T. and Griffin, B.E.
TITLE The EB virus genome in Daudi Burkitt's lymphoma cells has a
deletion similar to that observed in a non-transforming strain
(P3HR-1) of the virus
JOURNAL EMBO J. 3 (4), 813-821 (1984)
MEDLINE 84207939
PUBMED 6327290
REFERENCE
AUTHORS Biggin, M., Farrell, P.J. and Barrell, B.G.
TITLE Transcription and DNA sequence of the BamHI L fragment of B95-8
Epstein-Barr virus
JOURNAL EMBO J. 3 (5), 1083-1090 (1984)
MEDLINE 84236104
PUBMED 6203743
REFERENCE
AUTHORS Yates, J., Warren, N., Reisman, D. and Sugden, B.
TITLE A cis-acting element from the Epstein-Barr viral genome that
permits stable replication of recombinant plasmids in latently
infected cells
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)
MEDLINE 84222045
PUBMED 6328526
REFERENCE
AUTHORS Gibson, T., Stockwell, P., Ginsburg, M. and Barrell, B.
TITLE Homology between two EBV early genes and HSV ribonucleotide
reductase and 38K genes
JOURNAL Nucleic Acids Res. 12 (12), 5087-5099 (1984)
MEDLINE 84247360

```

PUBMED 6330697
 REFERENCE 14 (bases 1 to 171823)
 AUTHORS Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.V., Hattuli,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C., Tufnell,P.S. and Barrell,B.G.
 TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome
 JOURNAL Nature 310 (5974), 207-211 (1984)
 MEDLINE 84270667
 PUBMED 6087149
 REFERENCE 15
 AUTHORS Bodescot,M. and Perricaudet,M.
 TITLE Clustered alternative splice sites in Epstein-Barr virus RNAs
 JOURNAL Nucleic Acids Res. 15 (14), 5887 (1987)
 MEDLINE 87289053
 PUBMED 3039467
 REFERENCE 16
 AUTHORS Lauy,G., Perricaudet,M. and Farrell,P.J.
 TITLE A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome
 JOURNAL EMBO J. 7 (3), 769-774 (1988)
 MEDLINE 88283646
 PUBMED 2840285
 REFERENCE 17
 AUTHORS Parker,B.D., Bankier,A., Satchwell,S., Barrell,B. and Farrell,P.J.
 TITLE Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B95-8 deletion region
 JOURNAL Virology 179 (1), 339-346 (1990)
 MEDLINE 91021036
 PUBMED 2171209
 REFERENCE 18 (bases 1 to 171823)
 AUTHORS Hattuli,G.F., Barrell,B.G., Quinn,J. and McGeoch,D.
 JOURNAL Unpublished
 REFERENCE 19
 AUTHORS Binne,U.K., Amon,M. and Farrell,P.J.
 TITLE Induction of Epstein-Barr virus late promoters on small plasmids in the EBV late lytic cycle requires ori LYT
 JOURNAL Unpublished
 REFERENCE 20 (bases 1 to 171823)
 AUTHORS Farrell,P.J.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2002) Farrell P., Ludwig Institute for Cancer Research, Imperial College School of Medicine, St. Mary's Campus, Norfolk Place London W2 1PG
 COMMENT Construction:
 This sequence was assembled from B95-8 EBV [14] and Raji EBV [18] with sequence corrections [16,19]. The number of major internal repeat units has been reduced from 11.6 [14] to a more typical 7.6 and the B95-8 deletion sequences have been restored to give a sequence more representative of wild type EBV.
 Numbering
 Like the modified B95-8 sequence [14, 16] accession number V01555, this sequence starts 1 base to the left of the EcoRI site separating EcoRI DheI from EcoRI I (ie the first A of AAGATTC.).
 Location/Qualifiers
 1..171823
 /organism="Human herpesvirus 4"
 /mol_type="genomic DNA"
 /strain="B95-8"
 /db_xref="taxon:10376"
 join(1..139223,151555..171823)
 /organism="Human herpesvirus 4"
 /mol_type="genomic DNA"
 /strain="B95-8"
 /db_xref="taxon:10376"
 139223..151554
 /organism="Human herpesvirus 4"
 /mol_type="genomic DNA"
 /strain="Raj1"
 /db_xref="taxon:10376"
 join(166103..171823,1..16682)
 /gene="LMP2"
 join(166103..166458,58..272,360..458,540..788,871..951,

1026..1196,1280..1495,1574..1680)
 /gene="LMP2"
 /codon_start=1
 /product="terminal protein LMP2a"
 /protein_id="CAD53382.1"
 /db_xref="GI:23893577"
 /translation="MGSLEMPWAGAPSPGSDPDGDNNSQYPSASGSGSTPTP
 PNDREESNEBPPPEYEDPYWNGDGRSDYPLGTODOSLYIGLQHDNDGLPPPS
 PRDSSQHIYEBAGRSNMPVCLPVYVAIAAASCTAASCTVATGIALS
 LLIAAVASSYAAORKILTPVTLTAATFEPAICLWRIEDPPNSLFLAALAG
 LGIVLVNVLITLAVRRMRRLVCGGIMPLACVLVDVAVQLSPILGAVTVS
 MTLILATVPLMSRPGUCLTGAALLTLAALALASLIGLNTITFILMTLTV
 VLICSSCSCLPSKILARFLVALALLASALVAGSLIQTFKISSTFRLPNL
 FCMILIVAGLIFLIALITLWGSNGRTYGPVFCGLITVAGAVLTVASNTLLSA
 MLTAGFLIFLIGFALFVIRRCRCYCYCLTLESEBRRPPTFNTV"
 58..272
 /gene="LMP2"
 /number=2
 join(59..272,360..458,540..788,871..951,1026..1196,
 1280..1495,1574..1680)
 /gene="LMP2"
 /codon_start=1
 /product="terminal protein LMP2b"
 /protein_id="CAD53383.1"
 Query Match 100.0%; Score 20; DB 14; Length 171823;
 Best Local Similarity 100.0%; Pred. No. 0.027; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 0;
 CDS
 exon
 1 AGGGATGCTGACACAGA 20
 Db 36522 AGGGATGCTGACACAGA 36541
 RESULT 5
 EBV
 LOCUS 172281 bp DNA circular VRL 20-SEP-1999
 DEFINITION Epstein-Barr virus (EBV) genome, strain B95-8.
 ACCSSION V01555 J02070 K01730 V01554 X00498 X00499 X00784
 VERSION V01555.1 GI:59074
 KEYWORDS DNA polymerase; EBNA; genome; ribonucleotide reductase; tandem repeat; terminal repeat.
 SOURCE Human herpesvirus 4 (Epstein-Barr virus)
 ORGANISM Human herpesvirus 4
 Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Lymphocryptovirus.
 REFERENCE 1 (bases 1 to 172281)
 AUTHORS Arrand,R.R., Kyto,L., Walsh,J.E., Bjorck,E., Lindahl,T. and Griffin,B.E.
 TITLE Molecular cloning of the complete Epstein-Barr virus genome as a set of overlapping restriction endonuclease fragments
 JOURNAL Nucleic Acids Res. 9 (13), 2999-3014 (1981)
 MEDLINE 82014887
 PUBMED 6269068
 REFERENCE 2 (bases 1 to 172281)
 AUTHORS Kozak,M.
 TITLE Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes
 JOURNAL Nucleic Acids Res. 9 (20), 5233-5262 (1981)
 MEDLINE 82059504
 PUBMED 7301588
 REFERENCE 3 (bases 1 to 172281)
 AUTHORS Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.
 TITLE Sequence analysis of in vitro transcription of portions of the Epstein-Barr virus genome
 JOURNAL J. Cell. Biochem. 19 (3), 267-274 (1982)
 MEDLINE 83109311
 PUBMED 6296170
 REFERENCE 4 (bases 1 to 172281)
 AUTHORS Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.
 TITLE Homologous upstream sequences near Epstein-Barr virus promoters
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)
 MEDLINE 83169725

PUBMED 6300857
 REFERENCE 5 (bases 142687 to 159853)
 AUTHORS Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G.
 TITLE Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8
 JOURNAL Mol. Biol. Med. 1 (1), 21-45 (1983)
 MEDLINE 85035713
 PUBMED 6092825
 REFERENCE 6 (bases 112620 to 125316)
 AUTHORS Seguin,C., Farrell,P.J. and Barrell,B.G.
 TITLE DNA sequence and transcription of the BamHI fragment B region of B95-8 Epstein-Barr virus
 JOURNAL Mol. Biol. Med. 1 (3), 369-392 (1983)
 MEDLINE 85060424
 PUBMED 6094953
 REFERENCE 7 (bases 45644 to 52450)
 AUTHORS Jiang,K.T. and Hayward,S.D.
 TITLE Organization of the Epstein-Barr virus DNA molecule. III. Location of the P3HR-1 deletion junction and characterization of the NotI repeat units that form part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
 JOURNAL J. Virol. 48 (1), 135-148 (1983)
 MEDLINE 83294686
 PUBMED 6310141
 REFERENCE 8 (bases 159893 to 172281)
 AUTHORS Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J. and Barrell,B.G.
 TITLE DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences
 JOURNAL Mol. Biol. Med. 1 (4), 425-445 (1983)
 MEDLINE 85060428
 PUBMED 6094955
 REFERENCE 9 (bases 1 to 172281)
 AUTHORS Farrell,P.J., Bankier,A., Seguin,C., Deininger,P. and Barrell,B.G.
 TITLE Latent and lytic cycle promoters of Epstein-Barr virus
 JOURNAL EMBO J. 2 (8), 1331-1338 (1983)
 MEDLINE 20331131
 PUBMED 10872327
 REFERENCE 10 (bases 45415 to 52824)
 AUTHORS Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E.
 TITLE The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus
 JOURNAL EMBO J. 3 (4), 813-821 (1984)
 MEDLINE 84207939
 PUBMED 6327290
 REFERENCE 11 (bases 87650 to 92703)
 AUTHORS Biggin,M., Farrell,P.J. and Barrell,B.G.
 TITLE Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus
 JOURNAL EMBO J. 3 (5), 1083-1090 (1984)
 MEDLINE 84236104
 PUBMED 6203743
 REFERENCE 12 (bases 7315 to 9312)
 AUTHORS Yates,J., Warren,N., Reisman,D. and Sugden,B.
 TITLE A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)
 MEDLINE 84222045
 PUBMED 6328526
 REFERENCE 13 (bases 76089 to 79808)
 AUTHORS Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B.
 TITLE Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes
 JOURNAL Nucleic Acids Res. 12 (12), 5087-5099 (1984)
 MEDLINE 84247360
 PUBMED 6330697
 REFERENCE 14 (bases 1 to 172281)
 AUTHORS Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C., Tufnell,P.S. and Barrell,B.G.
 TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome

JOURNAL Nature 310 (5974), 207-211 (1984)
 MEDLINE 84270667
 PUBMED 6087149
 REFERENCE 15 (bases 1 to 172281)
 AUTHORS Boddeke,M. and Perricaudet,M.
 TITLE Clustered alternative splice sites in Epstein-Barr virus RNAs
 JOURNAL Nucleic Acids Res. 15 (14), 5887 (1987)
 MEDLINE 87289053
 PUBMED 3039467
 REFERENCE 16 (bases 1 to 172281)
 AUTHORS Laux,G., Perricaudet,M. and Farrell,P.J.
 TITLE A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome
 JOURNAL EMBO J. 7 (3), 769-774 (1988)
 MEDLINE 88283646
 PUBMED 2840285
 REFERENCE 17 (bases 1 to 172281)
 AUTHORS Hatfull,G.F., Barrell,B.G., Quinn,J. and McGeech,D.
 TITLE Unpublished
 JOURNAL 18 (bases 1 to 172281)
 REFERENCE 18 (bases 1 to 172281)
 AUTHORS Farrell,P.J. and Barrell,B.G.
 TITLE Direct Submission
 JOURNAL Submitted (05-JUN-1984)
 REFERENCE 19 (bases 1 to 172281)
 AUTHORS Farrell,P.J.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAR-1986) Farrell P., Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG

COMMENT

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been included which might represent exons or small frames because they occur in a logical combination with other genes because of some other experimental data. The reading frames are named according to the Bam HI fragment in which they start, eg BAPF3 is the third leftward frame starting in Bam HI fragment A. BOPF1 is the first rightward frame in Bam HI fragment O. If there is an obvious PATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITES OF POLYA SIGNALS
 This feature lists all occurrences of the sequence AATAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog AATAA is only listed when it is found in a position close to the end of a major reading frame.

SITES OF DONOR AND ACCEPTOR SEQUENCES
 This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case. Restriction enzyme SITES.
 Only the positions of the sites Bam HI (BAM) are listed.

RPT
 This feature is used to define repetitive sequences.

SITE DEL.
 This feature defines deletions in B95-8 with respect to other strains such as RAJ1 and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.

SITE HPN
 Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurrences noted.

ORIGIN
Denotes the region that encompasses an origin of replication (ori
F). (13).

NUMERING
The DNA sequence of B95-8 EBV has been revised [19]. The original
(Baer et al, 1984) base 359 has been deleted so the new sequence
around that position reads TCAGCTTT. To avoid renumbering the
entire sequence, position 1 has been moved 1 base to the left of
the EcoRI site separating EcoRI Duet from EcoRI I
(ie the first A of AGATTTC).

FEATURES
source
1..172281
/organism="Human herpesvirus 4"
/mol_type="genomic DNA"
/strain="B95-8"
/db_xref="taxon:10376"
58..272

MRNA
Query Match 100.0%; Score 20; DB 14; Length 172281;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGGACACAGA 20
Db 48810 AGGATGCTGGACACAGA 48829

RESULT 6
HSA8958RAJ 184113 bp DNA linear VRL 12-APR-1996
LOCUS Epstein-Barr virus, artificial joining of B95-8 complete genome
DEFINITION and the sequences from Raji of the large deletion found in B95-8.
ACCESSION M80517 M75389
VERSION M80517.1 GI:330330
KEYWORDS Human herpesvirus 4 (Epstein-Barr virus)
SOURCE Human herpesvirus 4
ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE
1 (sites)
Baer, R.J., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J.,
Gibson, T.J., Hatfull, G.F., Hudson, G.S., Satchwell, S.C., Segun, C.,
Tuffnell, P.S. and Barrett, B.G.
Nucleic Acids Res. 1984; 12(1): 51-58.
TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome
JOURNAL Nature 310 (5974), 207-211 (1984)
MEDLINE 84270667
PUBMED 6087149
REFERENCE
2 (sites)
Parker, B.D., Bankier, A., Satchwell, S., Barrett, B. and Farrell, P.J.
Sequence and transcription of Raji Epstein-Barr virus DNA spanning
the B95-8 deletion region
Virology 179 (1), 339-346 (1990)
JOURNAL 91021036
MEDLINE 2171209
PUBMED 2171209
REFERENCE
3 (sites)
Sample, J., Brooks, L., Sample, C., Young, L., Rowe, M., Gregory, C.,
Rickinson, A. and Kieff, E.
Restricted Epstein-Barr virus protein expression in Burkitt
lymphoma is due to a different Epstein-Barr nuclear antigen 1
transcriptional initiation site
Proc. Natl. Acad. Sci. U.S.A. 88 (14), 6343-6347 (1991)
JOURNAL 91296817
MEDLINE 1648738
PUBMED 1648738
REFERENCE
4 (bases 1 to 184113)
Jenson, H.B.
Genbank Curator Program
Unpublished (1992)
COMMENT
Original source text: Human herpesvirus 4 DNA.
The B95-8 genome (V01555) has a large deletion in the right side of
the genome which has been sequenced in Raji (M35547). These
sequences have been joined to form an extended and more complete,
although artifactual, EBV sequence.
For features, refer to feature tables of V01555 and M35547.

FEATURES
source
Location/Qualifiers
1..184113 "Human herpesvirus 4"
/organism="Human herpesvirus 4"
/mol_type="genomic DNA"
/db_xref="taxon:10376"
1..152008
/note="B95-8 sequences (corresponds to 1-152,008 of
V01555)."
152009..152012
/note="Overlap of B95-8 and Raji sequences at B95-8
deletion point (corresponds to 152,009-152,012 in V01555,
and 1-4 in M35547)."
153013..163839
/note="Raji sequences (corresponds to 5-11,831 of M35547)."
163840..163843
/note="Overlap of B95-8 and Raji sequences at B95-8
deletion point (corresponds to 152,009-152,012 of V01555,
and 11,832-11,835 of M35547)."
163844..184113
/note="B95-8 sequences (corresponds to 152,013-172,282 of
V01555)."
55824 a 55824 c 54622 g 37665 t

BASE COUNT 36002 a 55824 c 54622 g 37665 t

ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 184113;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTGGACACAGA 20
Db 48810 AGGATGCTGGACACAGA 48829

RESULT 7
AF272661 6076 bp mRNA linear ROD 11-SEP-2000
LOCUS Rattus norvegicus alpha 4 type V collagen mRNA, complete cds.
DEFINITION AF272661
ACCESSION AF272661.1 GI:8568093
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 6076)
Chernousov, M.A., Rothblum, K., Tyler, W.A., Stahl, R.C. and Carey, D.J.
Schwann cells synthesize type V collagen that contains a novel
alpha 4 chain. Molecular cloning, biochemical characterization, and
high affinity heparin binding of alpha 4(V) collagen
J. Biol. Chem. 275 (36), 28208-28215 (2000)
JOURNAL 20428740
MEDLINE 10852920
PUBMED 10852920
REFERENCE
2 (bases 1 to 6076)
Chernousov, M.A., Rothblum, K.N., Tyler, W.A., Stahl, R.C. and
Carey, D.J.
Direct Submission
Submitted (26-MAY-2000) Weis Center for Research, Penn State
College of Medicine, 100 N. Academy Ave., Danville, PA 17822, USA
Location/Qualifiers
1..6076
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague Dawley"
/db_xref="taxon:10116"
/cell_type="Schwann"
56..5269
/note="collagen alpha chain"
/codon_start=1
/product="alpha 4 type V collagen"
/protein_id="AA076432.1"
/db_xref="GI:8568093"
/translation="MRECRRLDQLQAGLCILLASLQVLSWTAAEPVVIETLGVHSD


```

*****
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
provided by the submittor.
This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.
1 240584: contig of 240584 bp in length.
Location/Qualifiers
1. 240584
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-137P24"
BASE COUNT 6117 a 57754 c 55691 g 57158 t 8864 others
ORIGIN
Query Match 95.0% Score 19; DB 2; Length 240584;
Best Local Similarity 100.0%; Pred No. 0.13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGATGCTCGACACAG 19
165989 AGGATGCTCGACACAG 165971
Db 165989 AGGATGCTCGACACAG 165971

RESULT 9
AC118348/c 254452 bp DNA linear HTG 21-SEP-2002
LOCUS Rattus norvegicus chromosome 18 clone CH230-221G18, *** SEQUENCING
DEFINITION IN PROGRESS ***
AC118348
AC118348.3 GI:23265687
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
VERSION Rattus norvegicus (Norway rat)
KEYWORDS Rattus norvegicus
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 254452)
Muzny, D., Maric, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, W., Baca, S., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Chen, Z., Chu, J.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
David, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Diaper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, T., Foster, P.,
Fraser, C. M., Gabris, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgescu, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, C.,
Henderson, R., Hines, S., Hladik, S., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulik, S., Hume, J., Idler, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Louie, H., Lozano, R., Lu, X., Ma, U.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McCreod, M. P., McNeill, T. Z., Meenen, E.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 254452)
Unpublished
Direct Submission
Submitted (17-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 254452)
Rat Genome Sequencing Consortium.
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence only
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- genome center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GNMO
Center clone name: CH230-221G18
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 232725 bases at least Q40
Consensus quality: 234773 bases at least Q30
Consensus quality: 235722 bases at least Q20
Estimated insert size: 256879; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
NOTE: This sequence may represent more than one clone.
NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
provided by the submittor.
This sequence will be replaced
by the finished sequence as soon as it is available and
the accession number will be preserved.
1 254452: contig of 254452 bp in length.
Location/Qualifiers

```

FEATURES


```

repeat_region 33747. .33871

Query Match      85.0%; Score 17; DB 9; Length 235357;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGGATGCTGGACACA 17
|||||
Db 78276 AGGATGCTGGACACA 78292

RESULT 12
AC095487 246186 bp DNA linear HTG 09-MAY-2003
LOCUS Rattus norvegicus clone CH230-7L4, WORKING DRAFT SEQUENCE.
DEFINITION AC095487 GI:30467639
ACCESSION AC095487
VERSION AC095487.6
KEYWORDS HTG; HTGS; PHASE2; HTGS; DRAFT; HTGS; FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 246186)
Munry,D,Marie, Metzker,M,Lee, Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsdrooks,S., Amin,A., Angiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,C., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., B Souza,L.,
Dayala,M.L., Davis,S., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Diaper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falle,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgescu,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamli,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howells,S., Hulak,J., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kover,C.,
Kowalski,C., Kraft,C.L., Lebow,H., Levan,D., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Louised,L., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mamhany,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,N., Mundaya,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,G., Olarnpungson,A., Pal,S., Paris,K.,
Nwackemehe,O., Okwou,G., Olarnpungson,A., Pal,S., Paris,K.,
Pasernak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poldexter,A., Popovic,D., Primus,E., Pu,L.,
Puato,M., Quirroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,M., Savary,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,D., Waldron,L., Walker,B., Wang,U.,
Wais,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,U.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wiczek,K., Woden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

```

```

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 246186)
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 246186)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:2491051.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a contig-scaffold). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GCLB
Center clone name: CH230-7L4
----- Summary Statistics
Assembly program: Atlas
Consensus quality: 232210 bases at least Q40
Consensus quality: 234785 bases at least Q30
Consensus quality: 236811 bases at least Q20
Estimated insert size: 245983; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_direct_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 246186: contig of 246186 bp in length.

Location/Qualifiers
1..246186 "Rattus norvegicus"
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-7L4"
1..1335
misc_feature 1..1335
note="wgs_contig"
BASE COUNT 63279 a 54327 c 54782 g 65902 t 7896 others
ORIGIN
Query Match      85.0%; Score 17; DB 2; Length 246186;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGGATGCTGGACACA 17
|||||
Db 177471 AGGATGCTGGACACA 177455

```

RESULT	13
AC095181	
LOCUS	
DEFINITION	AC095181 257316 bp DNA linear HTG 09-NOV-2002
ACCESSION	Rattus norvegicus clone CH230-9C15, *** SEQUENCING IN PROGRESS ***
VERSION	AC095181
KEYWORDS	3 unordered pieces.
SOURCE	AC095181.5 GI:24817891
ORGANISM	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED. Rattus norvegicus (Norway rat) Rattus norvegicus Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 257316)
AUTHORS	Muzny,D.Maile,,Metzker,M.Lee,,Abramzon,S.,Adams,C.,Alder,J.,

```

misc_feature      complement(250295..250549)
                  clone_end:sp6
                  site:ecori
                  end_sequence:BH339760"
misc_feature      252879..254160
                  /note="wgs end extension
                  clone_end:sp6"
BASE COUNT      66998 a 51599 c 49730 g 62175 t 24814 others
ORIGIN
Query Match      85.0%; Score 17; DB 2; Length 257316;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AGGATGCTGCTGACACA 17
|||||
Db      166282 AGGATGCTGCTGACACA 166298

RESULT 14
AC119530
LOCUS      Rattus norvegicus clone CH230-180D6, WORKING DRAFT SEQUENCE, 5
DEFINITION      Unordered pieces.
ACCESSION      AC119530
VERSION      AC119530.5 GI:25097886
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
              1 (bases 1 to 278519)
REFERENCE
  1 Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
    Allen,C., Allen,H., Alsbrooks,S., Amin,A., Arguliano,D.,
    Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
    Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
    Bistalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
    Bryant,N., Bunay,C., Burch,P., Burrell,K., Calderon,E.,
    Cardenas,V., Carter,K., Cavazos,I., Cesear,H., Center,A.,
    Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
    Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
    Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
    Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
    Drepper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
    Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
    Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
    Frazer,C.M., Gabisi,A., Gante,R., Garcia,A., Garner,T., Garza,M.,
    Gbagresorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
    Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
    Harvey,Y., Hawlak,P., Hawes,A., Henderson,N., Hernandez,J.,
    Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
    Hollins,B., Howells,S., Hulik,S., Hume,J., Idelbird,P., Jackson,A.,
    Jackson,S., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Joliver,A.,
    Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovac,C.,
    Kowals,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
    Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
    Lorenshewa,L., Louisedge,H., Lozada,R.J., Lu,X., Ma,J.,
    Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
    Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
    Mawliny,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
    Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
    Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L.,
    Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
    Nwaekeleme,O., Okwoum,G., Olarunpasegun,A., Palis,S., Parks,K.,
    Pasernak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
    Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
    Puata,M., Quintz,J., Rachelin,B., Reeves,K., Regier,M.A., Reich,R.,
    Reilly,B., Reilly,M., Ren,Y., Reuter,W., Richards,S., Riggs,F.,
    Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruit,S.J.,
    Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
    Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
  
```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Sytek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tinney,L., Trejos,Z., Usmani,K.,
Valae,R., Vera,V., Villasana,D., Walron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Wilson,R., Wiczysk,R., Woden,H., Wooley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., Von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submision
Unpublished
2 (bases 1 to 278519)
Worley,K.C.
Submitted (28-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 278519)
Rat Genome Sequencing Consortium.
Direct Submision
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23265331.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both ends sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center Project name: GVIS
Center clone name: CH230-180D6
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 229221 bases at least Q40
Consensus quality: 231985 bases at least Q30
Consensus quality: 233265 bases at least Q20
Estimated insert size: 234563; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 37221: contig of 37221 bp in length
* 37222: gap of unknown length
* 37223: contig of 209402 bp in length
* 37224: gap of unknown length
* 246823: gap of unknown length
* 246824: contig of 1925 bp in length
* 248748: gap of unknown length
* 248749: contig of 3845 bp in length
* 248849: gap of unknown length
* 252693: gap of unknown length
* 252793: contig of 25726 bp in length.
* 252794
* Location/Qualifiers
  
```

FEATURES

```

source
1. 278519
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-16CD6"
423.824
misc_feature
/clone="clone_boundary"
site:
clone_end:sp6
end_sequence:BH283438"
37332.38354
/misc_feature
/clone="wgs contig"
206906.207540
/misc_feature
/clone="clone_boundary"
site:
clone_end:T7
end_sequence:BH283437"
245746.246723
/misc_feature
/clone="wgs_end_extension"
clone_end:T7"

BASE COUNT 72233 a 45144 c 45892 g 70779 t 44471 others
ORIGIN

Query Match 85.0%; Score 17; DB 2; Length 278519;
Best Local Similarity 100.0%; Pred. No. 2,7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCGATGCTTGACACA 17
|||||
Db 184747 AGCGATGCTTGACACA 184763

RESULT 15
AC099161 297264 bp DNA linear HTG 10-MAY-2003
LOCUS Rattus norvegicus clone CH230-50G6, *** SEQUENCING IN PROGRESS ***,
DEFINITION 13 unordered pieces.
AC099161
AC099161.5 GI:30522089
HTG: HTG_PHASE1; HTG_DRAFT; HTG_ENRICHED.
VERSION Rattus norvegicus (Norway rat)
KEYWORDS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteiostomi;
SOURCE Eumariota; Rodentia; Sciurognathi; Muridae; Murinae;
ORGANISM Rattus.
1 (bases 1 to 297264)
Muzny,D,Marle, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alibrooks,S, Amin,A, Anguiano,D,
Aryalabechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,R, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A,
Chacko,J, Chavez,D, Chen,K, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Decker,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K,
Diaper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K,
Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Fraser,C,M, Gabisi,A, Gantar,R, Garcia,A, Garner,T, Garza,M,
Gebregeorgis,B, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W,
Gunnarsson,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K,
Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M,
Hollins,B, Howells,S, Huylk,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
Kwis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,U, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorenswater,L, Louieleged,H, Lozada,R,J, Lu,X, Ma,U,
Maheshwari,M, Mahindartine,M, Mahmoud,M, Mallory,K, Mangum,A,
Mangum,B, Mapa,P, Martin,K, Martin,R, Martinez,E,

```

```

TITLE JOURNAL
REFERENCE JOURNAL
AUTHORS Mawhiney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milesavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
Nwackemeah,O, Okwou,G, Olarpunaggon,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Plankoch,C,
Plapper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L,
Puzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J,
Sanders,W, Savery,G, Scherer,S, Scott,G, Shatman,S, Shen,H,
Shetty,J, Shvartsbeyn,A, Sison,I, Sitter,C,D, Smajls,D,
Sneed,A, Sodergren,E, Song,X,Z, Soreiller,R, Sosa,J,
Steinle,M, Strong,R, Sutton,A, Sytek,A, Taber,P, Taylor,C,
Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K,
Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J,
Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F,
Williams,G, Willson,R, Wiczysk,R, Wooder,H, Worley,K,
Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V,
Yu,S, Zhang,J, Zhou,U, Zhou,X, Zhao,S, Dunn,D, von
Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O,
Weinstock,G, and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 297264)
Mortley,K.C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 297264)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:23269124.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented. The sequence
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: G3ZH
Center clone name: CH230-50G6

----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 280761 bases at least Q40
Consensus quality: 28696 bases at least Q30
Estimated insert size: 31192; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	286994:	gap of 266934 bp in length
*	266935	267034:	gap of unknown length
*	267035	267034:	contig of 8990 bp in length
*	276025	276124:	gap of unknown length
*	276125	277611:	contig of 1487 bp in length
*	277612	277711:	gap of unknown length
*	277712	279271:	contig of 1560 bp in length
*	279272	279371:	gap of unknown length
*	279372	280408:	contig of 1037 bp in length
*	280409	280508:	gap of unknown length
*	280509	281665:	contig of 1154 bp in length
*	281665	281765:	gap of unknown length
*	281765	282455:	contig of 2783 bp in length
*	282456	284645:	gap of unknown length
*	284646	286178:	contig of 1533 bp in length
*	286179	286278:	gap of unknown length
*	286278	288037:	contig of 1759 bp in length
*	288037	288137:	gap of unknown length
*	288138	289477:	contig of 1340 bp in length
*	289478	289577:	gap of unknown length
*	289578	291208:	contig of 1631 bp in length
*	291209	291308:	gap of unknown length
*	291309	293349:	contig of 2185 bp in length
*	293349	293559:	gap of unknown length
*	293559	293766:	contig of 3671 bp in length.

FEATURES	Location/Qualifiers
source	1. .297264

```
misc_feature      265656.  .266934
```

BASE COUNT	87600	a	56301	c_57441	g	89645	t	6277	otheres
ORIGIN									

Query Match	85.0%;	Score 17;	DB 2;	Length 297264;
Best Local Similarity	100.0%;	Pred. No. 2.7;		
Matches 17;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	AGGATGCTGGACACA	17
Db	202860	AGGATGCTGGACACA	202876

LOCUS	AF542091	1377 bp	mRNA	linear	VRT 03-OCT-2002
RESULT 16	AF542091				

REFERENCE 1 (Pages 1 to 1177)
AUTHORS Gjoen T. and Berg T.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2002) Microbiology, School of Pharmacy,
University of Oslo, PO Box 1068 Blindern, Oslo 0316, Norway

CDS

```
/note="lipoprotein receptor"  
/codon_start=3
```

BASE COUNT	370	a	419	c	342	g	246	t
ORIGIN		ATTATLDP						
Query Match		80.0%;	Score 16;	DB 5;	Length 1377;			
Best Local Similarity		100.0%;	Pred. No. 10;					
Matches 16;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Oy	5	ATGCGCTGACACACAGA	20					
DB	102	ATGCGCTGACACACAGA	117					

BASE COUNT	370 a	419 c	342 g	246 t
ORIGIN				

```
Query Match      80.0%; Score16; DB 5; Length 1377;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Oy      5 ATGCCCTGGACACACAAGA 20
         |||||
Db     102 ATGCTTGACACACAAGA 117

```

LOCUS	AX705192	1452 bp	DNA	linear	PAT 04-APR-2003
DEFINITION	Sequence 3 from Patent WO03014358.				
ACCESSION	AX705192				
VERSION	AX705192.1				
KEYWORDS	GI:29561890				

REFERENCE
AUTHORS
TITLE
JOURNAL
1
Koltwitz, B., Breves, R. and Maurer, K. H.
Detergent and cleaning agent with hybrid alpha amyloses
Patent: WO 03014358-A 8 20-FEB-2003
Henkel Kommanditgesellschaft auf Aktien (DE)

CDS

BASE COUNT 426 a 291 c 388 g 347 t

BASE COUNT	426 a	291 c	388 g	347 t
ORIGIN				
Query Match		80.0%;	Score 16;	DB 6;
Best Local Similarity		100.0%;	Pred. No. 10;	
Matches	16;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0

Query Match	80.0%	Score 16	DB 6	Length 1452
Best Local Similarity	100.0%	Pred. No. 10		
Matches 16	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	3	GGATGCTGGACACAA	18	
Db	950	GGATGCTGGACACAA	935	

QY 3 GGATGCCCTGGACACAA 18
|||
Db 950 GGATGCCCTGGACACAA 935

RESULT 18
AX705202/C

LOCUS AX705202 1452 bp DNA linear PAT 04-APR-2003
DEFINITION Sequence 13 from Patent WO03014358.
ACCESSION AX705202
VERSION AX705202.1 GI:29561900
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial sequences.
REFERENCE
1
Kotwitz,B., Breves,R. and Maurer,K.H.
Detergent and cleaning agent with hybrid alpha amylases
Patent: WO 03014358-A 13 20-FEB-2003;
Henkel Kommanditgesellschaft auf Aktien (DE)
Location/Qualifiers
1. 1452
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="Fusion der Alpha-Amylase-Gene von B. licheniformis
und B. amyloliquefaciens (ALM34-84)."
<1. 1452
/note="unmated protein product"
/codon_start=1
/protein_table=1
/transl_table=11
/translation="VNGTLMQYFEWYTPNDGQHKRLQNDAEHLSDIGITAVWIPPAY
KGTSDQVGIAGAYDLVDFGHQGTFRVYTGKSLQALSGLSHRNVQVGDVVLN
HKAGADATIEDVAVNPANRNQDTESEVQIKAMTDFRPGKNTYSDFKMWHYHG
ADDESKISRIKFRPGEKAMDEYSENGVNDYIMADVYDHPDVAAETKNGIM
YANELSLDFGRIDAALHIFKPSFLRWVQAVROATGKEMFTVAEYVQNNAGKLENTLNK
TSFNQSVFDPVPLHFNLOAASQGGVDMRLDGTAVSHRPEKAVTFVFNHDTQGS
LESTVQVPEKLVAVTILPREGSPQVFGYDMVGTGTSPEKIPSLKDIIEFLILARK
EYAVGRQHDYIDHPVIGWTRREGSDSSAAKGLAALITDPPGSKRYTAGLKNAGETWY
DITGNSDYTKISDGEHFVNDSSVSIYQK"

BASE COUNT 416 a 290 c 398 g 348 t
ORIGIN
Query Match 80.0%; Score 16; DB 6; Length 1452;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 950 GGATGCTGGACACAA 935

RESULT 19
LOCUS AX705204 1458 bp DNA linear PAT 04-APR-2003
DEFINITION Sequence 15 from Patent WO03014358.
ACCESSION AX705204
VERSION AX705204.1 GI:29561902
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial sequences.
REFERENCE
1
Kotwitz,B., Breves,R. and Maurer,K.H.
Detergent and cleaning agent with hybrid alpha amylases
Patent: WO 03014358-A 15 20-FEB-2003;
Henkel Kommanditgesellschaft auf Aktien (DE)
Location/Qualifiers
1. 1458
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/note="Fusion der Alpha-Amylase-Gene von B. licheniformis
und B. amyloliquefaciens (LAL19-433)."
<1. 1458
/note="unmated protein product"
/codon_start=1
/transl_table=11

CDS
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
ORIGIN
Query Match
Best Local Similarity
Matches

/protein_id="CAD88118.1"
/db_xref="GI:29561903"
/translation="ANNGTLMQYFEWYTPNDGQHKRLQNDAEHLSDIGITAVWIPP
AYKSGSDNGVGYDYDLGFOQGTFRVYTGKSLQALSGLSHRNVQVGDVVLN
HKAGADATIEDVAVNPANRNQDTESEVQIKAMTDFRPGKNTYSDFKMWHYHG
DGAHMDSEKTSRIKFRPGEKAMDEYSENGVNDYIMADVYDHPDVAAETKNGIM
IYANELSLDFGRIDAALHIFKPSFLRWVQAVROATGKEMFTVAEYVQNNAGKLENTLNK
TSFNQSVFDPVPLHFNLOAASQGGVDMRLDGTAVSHRPEKAVTFVFNHDTQGS
QLESTVQVPEKLVAVTILPREGSPQVFGYDMVGTGTSPEKIPSLKDIIEFLILARK
RKEYAVGRQHDYIDHPVIGWTRREGSDSSAAKGLAALITDPPGSKRYTAGLKNAGET
WYDITGNSDYTKISDGEHFVNDSSVSIYQK"

BASE COUNT 422 a 292 c 395 g 349 t
ORIGIN
Query Match 80.0%; Score 16; DB 6; Length 1458;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 956 GGATGCTGGACACAA 941

RESULT 20
LOCUS E51824 1545 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel amylase for baking and gene thereof.
ACCESSION E51824
VERSION E51824.1 GI:18628294
KEYWORDS JP 2000135093-A/1.
SOURCE
ORGANISM
unidentified
unclassified
unidentified
unclassified
REFERENCE
1 (bases 1 to 1545)
Iamagawa,S., Yoshida,M., Minoda,M., Takahashi,S., Hidaki,Y.,
Tani,M. and Hashimoto,S.
Novel amylase for baking and gene thereof
Patent: JP 2000135093-A 1 16-MAY-2000;
JOURNAL
DAIWA KASEI KK
OS Bacillus amyloliquefaciens, clone No. 21
PN JP 2000135093-A/1
PD 16-May-2000
PF 20-AUG-1999 JP 1999234813
PR
PI SHINICHIRO TAMAGAWA,MASAHARU YOSHIDA,MASASHI MINODA, PI
SATOKO TAKAHASHI,
PI YUMIKO HIDAKI,MASAKAZU TANI,SATOSHI HASHIMOTO PC
CI2N15/09,A21D2/24,CI2N1/21,CI2N9/28//CI2N15/09,CI2R1/07, PC
(CI2N1/21,CI2R1:125), (CI2N9/28,CI2R1:125), (CI2N15/09, CI2N15/00, PC
CI2R1:07)
CC
FH Key Location/Qualifiers
FT CDS (1)..(1545)
FT sig_peptide (1)..(93)
FT mat_peptide (94)..(1542).
FEATURES
source
1. 1545
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 454 a 311 c 406 g 374 t
ORIGIN
Query Match 80.0%; Score 16; DB 6; Length 1545;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1043 GGATGCTGGACACAA 1028

RESULT 21

ES1825/c
LOCUS ES1825 1545 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel amylase for baking and gene thereof.
ACCESSION ES1825
VERSION ES1825.1 GI:18628295
KEYWORDS JP 2000135093-A/2.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1545)
AUTHORS Tamagawa,S., Yoshida,M., Minoda,M., Takahashi,S., Hidaki,Y.,
Tani,M. and Hashimoto,S.
TITLE Novel amylase for baking and gene thereof
JOURNAL Patent: JP 2000135093-A 2 16-MAY-2000;
DAIWA KASEI KK
OS Bacillus amyloliquefaciens, No. 22
PN JP 2000135093-A/2
PD 16-MAY-2000
PF 20-AUG-1999 JP 1999234813
PR SHINICHIRO TAMAGAWA, MASAHARU YOSHIDA, MASASHI MINODA, PI
SATOKO TAKAHASHI,
PI YUMIKO HIDAKI, MASAKAZU TANI, SATOSHI HASHIMOTO PC
C12N15/09, A21D2/24, C12N1/21, C12N9/28, C12N15/09, C12R1:07), PC
(C12N1/21, C12R1:125), (C12N9/28, C12R1:125), C12N15/00, (C12N15/00, PC
C12R1:07)
CC
FH Key Location/Qualifiers
FT CDS (1)..(1545)
FT sig_peptide (1)..(93)
FT mat_peptide (94)..(1542).
LOCATION/Qualifiers
1..1545
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 454 a 309 c 406 g 376 t
ORIGIN
Query Match 80.0%; Score 16; DB 6; Length 1545;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGGACACAA 18
DB 1043 GGATGCTGGACACAA 1028
RESULT 22
LOCUS ES1826 1545 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel amylase for baking and gene thereof.
ACCESSION ES1826
VERSION ES1826.1 GI:18628296
KEYWORDS JP 2000135093-A/3.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1545)
AUTHORS Tamagawa,S., Yoshida,M., Minoda,M., Takahashi,S., Hidaki,Y.,
Tani,M. and Hashimoto,S.
TITLE Novel amylase for baking and gene thereof
JOURNAL Patent: JP 2000135093-A 3 16-MAY-2000;
DAIWA KASEI KK
OS Bacillus amyloliquefaciens, No. 24
PN JP 2000135093-A/3
PD 16-MAY-2000
PF 20-AUG-1999 JP 1999234813
PR SHINICHIRO TAMAGAWA, MASAHARU YOSHIDA, MASASHI MINODA, PI
SATOKO TAKAHASHI,
PI YUMIKO HIDAKI, MASAKAZU TANI, SATOSHI HASHIMOTO PC
C12N15/09, A21D2/24, C12N1/21, C12N9/28, C12N15/09, C12R1:07), PC

(C12N1/21, C12R1:125), (C12N9/28, C12R1:125), C12N15/00, (C12N15/00, PC
C12R1:07)
CC
FH Key Location/Qualifiers
FT CDS (1)..(1545)
FT sig_peptide (1)..(93)
FT mat_peptide (94)..(1542).
LOCATION/Qualifiers
1..1545
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 454 a 309 c 406 g 376 t
ORIGIN
Query Match 80.0%; Score 16; DB 6; Length 1545;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGGACACAA 18
DB 1043 GGATGCTGGACACAA 1028
RESULT 23
LOCUS ES1827/c 1545 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel amylase for baking and gene thereof.
ACCESSION ES1827
VERSION ES1827.1 GI:18628297
KEYWORDS JP 2000135093-A/4.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1545)
AUTHORS Tamagawa,S., Yoshida,M., Minoda,M., Takahashi,S., Hidaki,Y.,
Tani,M. and Hashimoto,S.
TITLE Novel amylase for baking and gene thereof
JOURNAL Patent: JP 2000135093-A 4 16-MAY-2000;
DAIWA KASEI KK
OS Bacillus amyloliquefaciens, No. 25
PN JP 2000135093-A/4
PD 16-MAY-2000
PF 20-AUG-1999 JP 1999234813
PR SHINICHIRO TAMAGAWA, MASAHARU YOSHIDA, MASASHI MINODA, PI
SATOKO TAKAHASHI,
PI YUMIKO HIDAKI, MASAKAZU TANI, SATOSHI HASHIMOTO PC
C12N15/09, A21D2/24, C12N1/21, C12N9/28, C12R1:125), C12N15/00, (C12N15/00, PC
C12R1:07)
CC
FH Key Location/Qualifiers
FT CDS (1)..(1545)
FT sig_peptide (1)..(93)
FT mat_peptide (94)..(1542).
LOCATION/Qualifiers
1..1545
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 454 a 309 c 406 g 376 t
ORIGIN
Query Match 80.0%; Score 16; DB 6; Length 1545;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGGACACAA 18
DB 1043 GGATGCTGGACACAA 1028

RESULT 24
LOCUS A20154 1972 bp DNA linear PAT 18-AUG-1994
DEFINITION alpha-amyase gene (and flanking sequence).
ACCESSION A20154
VERSION A20154.1 GI:580682
KEYWORDS
SOURCE Bacillus amyloliquefaciens
ORGANISM Bacillus amyloliquefaciens
REFERENCE 1 (bases 1 to 1972)
AUTHORS Van Eijk, J.H., Quax, W.J. and Sanders, J.P.M.
TITLE Mutant enzyme having reduced stability under industrial application conditions
JOURNAL Patent: EP 0409299-A 4 23-JAN-1991;
GIST-BROCADES N.V.
FEATURES
source Location/Qualifiers
1..1972
/organism="Bacillus amyloliquefaciens"
/mol_type="genomic DNA"
/db_xref="taxon:1390"
138..1662
/codon_start=1
/transl_table=1
/product="alpha-amyase"
/protein_id="CAA01489.1"
/db_xref="GI:580683"
/db_xref="SWISS-PROT:P00692"
/translation="MIQKRRVSRFLVLMCTLLFVSLPIRTSAVNGTLMQYFEWYT
PNDGGHWRLOQDAEHLSDIGITAWIPAYKGLSOSDNGYEPYDLIDGFEQOKT
RTYGTGKSELQDAIGLSHRNVQVGVNLNKAAGADATEDVTAVERNANNOETS
EYQIKAMTDFRFRGNTSYDPKMHVDFGADWDSEKRSIRFKFRGGKAMDEVS
SENGNVDYLMADVDDVDPVAAETKMGIVANELSLDGPRIIDAHGIRSPIDMY
QAVROATGKEMFTVAEYNNONNAGKLENTANTSPNOSFEDVPLFPIQAASSOGGCGYD
MRRLDGTIVSRHPRKAVTFVNHDTOPQGSLESTVQWFKPLAFLITRESGYPV
FYGDMYGTGKSPKPEISLKNIEPIILKAREYVGPQHDYIDHPDVIQWTRGSSA
AKSGLAALITDGPSSKRMVAGLNKAGETWYDITGNRSDTVKISGDGMEFHVNDGSV
SIYQK"
BASE COUNT 565 a 398 c 518 g 491 t
ORIGIN
Query Match 80.0%; Score 16; DB 6; Length 1972;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGACACAA 18
DB 1180 GGATGCTGACACAA 1165

RESULT 25
LOCUS BACAM/c 2084 bp mRNA linear BCT 26-APR-1993
DEFINITION Bacillus amyloliquefaciens alpha-amyase gene, complete cds.
ACCESSION J01542 J01543 M12033 M12034
VERSION J01542.1 GI:142428
KEYWORDS alpha-amyase; amyase.
SOURCE Bacillus amyloliquefaciens
ORGANISM Bacillus amyloliquefaciens
REFERENCE 1 (bases 1 to 537)
AUTHORS Palva, I., Petersson, R.F., Kalkinen, N., Lehtovaara, P., Sarvas, M.,
Soderlund, H., Takkinen, K. and Kaariainen, L.
TITLE Nucleotide sequence of the promoter and NH2-terminal signal peptide
region of the alpha-amyase gene from Bacillus amyloliquefaciens
JOURNAL Gene 15 (1), 43-51 (1981)
MEDLINE 82051296
PUBMED 6170539
2 (bases 76 to 2084)
REFERENCE Takkinen, K., Petersson, R.F., Kalkinen, N., Palva, I., Soderlund, H.
AUTHORS and Kaariainen, L.
TITLE Amino acid sequence of alpha-amyase from Bacillus
JOURNAL amyloliquefaciens deduced from the nucleotide sequence of the

JOURNAL
MEDLINE 83109808
PUBMED 6185474
REFERENCE 3 (bases 182 to 237; 1796 to 1857)
AUTHORS Lehtovaara, P., Ulmanen, I. and Palva, I.
TITLE In vivo transcription initiation and termination sites of an
alpha-amyase gene from Bacillus amyloliquefaciens cloned in
Bacillus subtilis
JOURNAL Gene 30 (1-3), 11-16 (1984)
MEDLINE 85077601
PUBMED 6210229
COMMENT Original source text: Bacillus amyloliquefaciens Marburg strain-1h,
cDNA to mRNA.
The deduced amino acid sequence deviates slightly from a published
sequence for this species (four conflicts). The signal peptide
seems unusually large (31 residues) compared to other known signal
peptides; the promoter and mRNA start point are speculatively
addressed.
FEATURES
source Location/Qualifiers
1..2084
/organism="Bacillus amyloliquefaciens"
/mol_type="mRNA"
/db_xref="taxon:1390"
220..1845
/product="alpha-amyase mRNA"
250..1794
/note="alpha-amyase protein precursor (EC 3.2.1.1)"
/codon_start=1
/transl_table=1
/protein_id="AAA22191.1"
/db_xref="GI:142428"
/translation="MIQKRRVSRFLVLMCTLLFVSLPIRTSAVNGTLMQYFEWYT
PNDGGHWRLOQDAEHLSDIGITAWIPAYKGLSOSDNGYEPYDLIDGFEQOKT
RTYGTGKSELQDAIGLSHRNVQVGVNLNKAAGADATEDVTAVERNANNOETS
EYQIKAMTDFRFRGNTSYDPKMHVDFGADWDSEKRSIRFKFRGGKAMDEVS
SENGNVDYLMADVDDVDPVAAETKMGIVANELSLDGPRIIDAHGIRSPIDMY
QAVROATGKEMFTVAEYNNONNAGKLENTANTSPNOSFEDVPLFPIQAASSOGGCGYD
MRRLDGTIVSRHPRKAVTFVNHDTOPQGSLESTVQWFKPLAFLITRESGYPV
FYGDMYGTGKSPKPEISLKNIEPIILKAREYVGPQHDYIDHPDVIQWTRGSSA
AKSGLAALITDGPSSKRMVAGLNKAGETWYDITGNRSDTVKISGDGMEFHVNDGSV
SIYQK"
BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN
Query Match 80.0%; Score 16; DB 1; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGATGCTGACACAA 18
DB 1292 GGATGCTGACACAA 1277

RESULT 26
LOCUS AR008285/c 2084 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 3 from patent US 5753460.
ACCESSION AR008285
VERSION AR008285.1 GI:3667394
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2084)
AUTHORS Bisgard-Frantzen, H., Borchert, T., Vedel, S., Svendsen, A., Thellertsen, M.
TITLE Amylase variants
JOURNAL Patent: US 5753460-A 3 19-MAY-1998;

FEATURES Location/Qualifiers
source 1..2084
/organism="unknown"
BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 27 AR037275 2084 bp DNA linear PAT 29-SEP-1999
LOCUS AR037275/c
DEFINITION Sequence 3 from patent US 5801043.
ACCESSION AR037275
VERSION AR037275.1 GI:5955131
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2084)
AUTHORS Bisg.ang.rd-Frantzen,H., Borchert,T.Vedel., Svendsen,A.,
Thelersen,M. and Van der Zee,P.
TITLE Amylase variants
JOURNAL Patent: US 5801043-A 3 01-SEP-1998;
FEATURES Location/Qualifiers
source 1..2084
/organism="unknown"
BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 28 AR052145 2084 bp DNA linear PAT 29-SEP-1999
LOCUS AR052145/c
DEFINITION Sequence 3 from patent US 5830837.
ACCESSION AR052145
VERSION AR052145.1 GI:5975509
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2084)
AUTHORS Bisg.ang.rd-Frantzen,H., Borchert,T.Vedel., Svendsen,A.,
Thelersen,M. and Van der Zee,P.
TITLE Amylase variants
JOURNAL Patent: US 5830837-A 3 03-NOV-1998;
FEATURES Location/Qualifiers
source 1..2084
/organism="unknown"
BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 29 AR087551 2084 bp DNA linear PAT 07-SEP-2000
LOCUS AR087551/c
DEFINITION Sequence 3 from patent US 5989169.
ACCESSION AR087551
VERSION AR087551.1 GI:10014314
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2084)
AUTHORS Svendsen,A., Bisg.ang.rd-Frantzen,H. and Borchert,T.Vedel.,
TITLE .alpha.-amylase mutants
JOURNAL Patent: US 5989169-A 3 23-NOV-1999;
FEATURES Location/Qualifiers
source 1..2084
/organism="unknown"
BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 30 AR129916 2084 bp DNA linear PAT 16-MAY-2001
LOCUS AR129916/c
DEFINITION Sequence 13 from patent US 6187576.
ACCESSION AR129916
VERSION AR129916.1 GI:14117813
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2084)
AUTHORS Svendsen,A., Borchert,T.Vedel. and Bisg.ang.rd-Frantzen,H.
TITLE .alpha.-amylase mutants
JOURNAL Patent: US 6187576-A 13 13-FEB-2001;
FEATURES Location/Qualifiers
source 1..2084
/organism="unknown"
BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 31 AR137905 2084 bp DNA linear PAT 16-JUN-2001
LOCUS AR137905/c
DEFINITION Sequence 31 from patent US 6197565.
ACCESSION AR137905
VERSION AR137905.1 GI:14479414
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2084)
AUTHORS Svendsen,A., Kjaerulff,S., Bisgaard-Frantzen,H. and Andersen,C.
TITLE .alpha.-Amylase variants

JOURNAL Patent: US 6197565-A 31-06-MAR-2001;
FEATURES Location/Qualifiers
SOURCE 1..2084
/organism="unknown"
BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 32
LOCUS AR143262/c 2084 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 58 from patent US 6204232.
ACCESSION AR143262
VERSION AR143262.1 GI:15104548
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2084)
AUTHORS Borchert,T.Vedel., Svendsen,A., Andersen,C., Nielsen,B.,
Nissen,T.,Lauesgaard, and Kj.ae.buted.rulff.sslashedren.
TITLE .alpha.-amylase mutants
JOURNAL Patent: US 6204232-A 58-20-MAR-2001;
FEATURES Location/Qualifiers
SOURCE 1..2084
/organism="unknown"

BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 33
LOCUS AR224261/c 2084 bp DNA linear PAT 26-SEP-2002
DEFINITION Sequence 3 from patent US 6440716.
ACCESSION AR224261
VERSION AR224261.1 GI:23333028
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2084)
AUTHORS Svendsen,A., Bisgard-Frantzen,H. and Borchert,T.V.
TITLE .alpha.-amylase mutants
JOURNAL Patent: US 6440716-A 3-27-AUG-2002;
FEATURES Location/Qualifiers
SOURCE 1..2084
/organism="unknown"

BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 34
LOCUS AR285306/c 2084 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 13 from patent US 6528298.
ACCESSION AR285306
VERSION AR285306.1 GI:29722408
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2084)
AUTHORS Svendsen,A., Borchert,T.V., Bisgard-Frantzen,H., Outtrup,H.,
Nielsen,B.R., Nielsen,V.S. and Hedegaard,L.
TITLE .alpha.-amylase mutants
JOURNAL Patent: US 6528298-A 13-04-MAR-2003;
FEATURES Location/Qualifiers
SOURCE 1..2084
/organism="unknown"

BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGATGCTGGACACAA 18
|||||
Db 1292 GGATGCTGGACACAA 1277

RESULT 35
LOCUS AX244195/c 2084 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 9 from Patent WO0166712.
ACCESSION AX244195
VERSION AX244195.1 GI:15859250
KEYWORDS
SOURCE Bacillus amyloliquefaciens
ORGANISM Bacillus amyloliquefaciens
REFERENCE 1
AUTHORS Andersen,C., Borchert,T.V. and Nielsen,B.R.
TITLE Variants with altered properties
JOURNAL Patent: WO 0166712-A 9-13-SEP-2001;
FEATURES Location/Qualifiers
SOURCE 1..2084
/organism="Bacillus amyloliquefaciens"
/mol_type="genomic DNA"
/db_xref="taxon:1390"
/note="BAN"
/codon_start=1
/transl_table=1
/protein_id="NC088551.1"
/db_xref="GI:15859251"
/translation="VNGTLMQYFPEWYTPDQGHKRLQNDADHLSDIGTAVWIPAY
KGLSQSDNGYGPYDLVDFQKGVTRKRYTKSLQDAIGSLHSRNVQVGDVVLN
HKAGADATEDVAVEVNPANRQDETSEEQIRAWDFRPPGGNTYSDKHWYHFDG
ADWDESRKISRIFKFRGEGKAMDEWSENGVYDLYADVDYDHPDVAETKMGIM
YANELSLDGFRIIDAKHKEFSLRDVQVROATGKEMFTVAEWMQNKLENYLNK
TSFNQSVFVPPVPHENIQASQGGVQVDRILDGTGVSSHPRKAVTFVNNHDPQGS
LESTQGTWFKRPLAYAFILTRSGYQVRYGDMYGRKTSKPKRISIKNIEPIIKRK
EYAYGQPHYIHPDVICWTRFGDSASAGLALITTDGPGSKRWYAGLKNAGETWY
DITGNSDVTVKIGSDGWEFFHVDGVSIVYVK"

BASE COUNT 610 a 401 c 544 g 529 t
ORIGIN

Query Match 80.0%; Score 16; DB 6; Length 2084;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Mon Aug 18 10:30:09 2003

us-10-074-620-5.oli.rge

Page 22

9

Oy 3 GGATGCTTGAACACAA 18
|||
Db 1292 GGATGCTTGAACACAA 1277
|||

Search completed: August 15, 2003, 09:34:04
uob time : 555.75 secs

Mon Aug 18 10:30:11 2003

us-10-074-620-6.011.rst

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:41:37 ; Search time 1377.2 Seconds
(without alignments)
388.250 Million cell updates/sec

Title: US-10-074-620-6

Perfect score: 22

Sequence: 1 atgccaccaccagcagcaccacca 22

Scoring table: Oligo_NUC
Gapco 60.0 , Gapext 60.0

Searched: 22781392 segs, 12152238056 residues

Word size : 10

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 120 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estmu:*
4: em_estnu:*
5: em_estgv:*
6: em_estbl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rnd:*
26: em_ges_phg:*
27: em_ges_vrt:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	22	100.0	900	13	BUS94332	AGENCOURT	BUS94332 AGENCOURT
2	20	90.9	516	14	CB447869	NA	CB447869 701906 NA
3	20	90.9	709	14	CB447515	NA	CB447515 701522 NA
4	19	86.4	301	12	BI675164	CA01	BI675164 949079CA01

C	5	19	86.4	370	12	BI596214	BI596214 949079CA01
C	6	19	86.4	586	28	AO158214	AO158214 nbx00010H
C	7	19	86.4	606	28	AO290064	AO290064 nbx00036C
C	8	19	86.4	627	28	AO542863	AO542863 RPT-11-3
C	9	19	86.4	629	12	BI064968	BI064968 Ppfin-PKO
C	10	19	86.4	1227	10	BF538979	BF538979 602048525
C	11	18	81.8	34	28	A2761910	A2761910 1M0556C05
C	12	18	81.8	118	12	BM096787	BM096787 EBma07 SQ
C	13	18	81.8	207	12	BI019327	BI019327 MR3-MT032
C	14	18	81.8	328	12	BM376718	BM376718 EBem05 SQ
C	15	18	81.8	324	13	BO882962	BO882962 OCE317NT1
C	16	18	81.8	321	10	BE398636	BE398636 WHE0231D
C	17	18	81.8	366	13	BO594281	BO594281 E012758-0
C	18	18	81.8	339	12	BI336904	BI336904 AR083G07S
C	19	18	81.8	396	13	BO593667	BO593667 E012763-0
C	20	18	81.8	399	14	C84299	C84299 C84299 oste
C	21	18	81.8	406	28	A2859178	A2859178 2M0164L07
C	22	18	81.8	410	12	BI596215	BI596215 949079CA01
C	23	18	81.8	425	13	BO593650	BO593650 E012763-0
C	24	18	81.8	425	14	CA123598	CA123598 Wdrlf-PKO
C	25	18	81.8	436	28	AO042664	AO042664 CIT-HSP-2
C	26	18	81.8	437	13	BO982501	BO982501 QGE16J21
C	27	18	81.8	443	28	A2234554	A2234554 RPT-23-5
C	28	18	81.8	447	13	BO593660	BO593660 E012763-0
C	29	18	81.8	474	13	BO395258	BO395258 NISC ng14
C	30	18	81.8	495	13	BO796727	BO796727 EST 5665
C	31	18	81.8	498	29	CNS074T1	AL611479 Anophelies
C	32	18	81.8	521	28	BH006015	BH006015 BMBAAC10U0
C	33	18	81.8	527	13	BO593655	BO593655 E012763-0
C	34	18	81.8	529	13	BO282622	BO282622 WHE3074 F
C	35	18	81.8	534	13	BO593645	BO593645 E012763-0
C	36	18	81.8	536	13	BO593658	BO593658 E012763-0
C	37	18	81.8	542	13	BO764923	BO764923 EBC801 SQ
C	38	18	81.8	552	14	CB517354	CB517354 S8A1L9D53
C	39	18	81.8	554	14	BO594283	BO594283 E012758-0
C	40	18	81.8	558	14	CD553292	CD553292 B0357F08-0
C	41	18	81.8	574	9	AW140795	AW140795 EST290790
C	42	18	81.8	584	14	CB458200	CB458200 716347 MA
C	43	18	81.8	589	28	A2286149	A2286149 RPT-23-1
C	44	18	81.8	601	10	BE492011	BE492011 WHE0555 G
C	45	18	81.8	608	13	BU119631	BU119631 603144164
C	46	18	81.8	611	12	BI723654	BI723654 1031067D1
C	47	18	81.8	611	13	BO406266	BO406266 GA_Ed009
C	48	18	81.8	620	10	BF251540	BF251540 EST418801
C	49	18	81.8	623	14	CB922302	CB922302 VPD091A09
C	50	18	81.8	629	13	BO402958	BO402958 GA_Ed005
C	51	18	81.8	629	13	BO402959	BO402959 GA_Ed005
C	52	18	81.8	633	10	BE291320	BE291320 WHE2221 G
C	53	18	81.8	635	28	A2282247	A2282247 RPT-23-1
C	54	18	81.8	665	14	CA486595	CA486595 WHE3433 E
C	55	18	81.8	666	13	BO870895	BO870895 QG10FL3
C	56	18	81.8	669	13	BO823252	BO823252 1030107D0
C	57	18	81.8	670	12	BI723652	BI723652 1031067D1
C	58	18	81.8	675	9	AW448832	AW448832 BRV 1589
C	59	18	81.8	675	13	BO605990	BO605990 BRV 1589
C	60	18	81.8	680	14	CA817536	CA817536 CA17E1204
C	61	18	81.8	680	14	CD491143	CD491143 WHE3074 F
C	62	18	81.8	689	28	A2866439	A2866439 2M0176M17
C	63	18	81.8	704	13	BO805152	BO805152 WHE3563 D
C	64	18	81.8	726	13	BU378193	BU378193 603588388
C	65	18	81.8	759	29	CNS021UX	AL177378 Tetraodon
C	66	18	81.8	781	13	BU353011	BU353011 603527258
C	67	18	81.8	841	13	BO430562	BO430562 AGENCOURT
C	68	18	81.8	869	13	BO420612	BO420612 AGENCOURT
C	69	18	81.8	888	13	BO422856	BO422856 AGENCOURT
C	70	18	81.8	899	29	CNS042NR	AL271128 Tetraodon
C	71	18	81.8	999	29	CNS03BES	AL236413 Tetraodon
C	72	18	81.8	1047	29	CNS05RCS	AL250366 Tetraodon
C	73	18	81.8	1155	12	BM803512	BM803512 AGENCOURT
C	74	18	81.8	1640	29	AC073459	AC073459 Pan trogl
C	75	17	77.3	81	14	N49052	N49052 yy78a11 r1
C	76	17	77.3	94	28	A2791401	A2791401 2M0041K16
C	77	17	77.3	132	10	BE125086	BE125086 Dg1_16_D0

C 78	17	77.3	161	28	A2592710	
C 79	17	77.3	164	14	CB116339	
C 80	17	77.3	168	10	BE598224	
C 81	17	77.3	184	9	AM094426	
C 82	17	77.3	185	9	AM214850	
C 83	17	77.3	190	26	B46974	
C 84	17	77.3	196	14	CD204318	
C 85	17	77.3	208	14	BB304836	
C 86	17	77.3	226	14	CA603727	
C 87	17	77.3	226	28	AZ837781	
C 88	17	77.3	232	10	BF803955	
C 89	17	77.3	241	10	BB595915	
C 90	17	77.3	253	9	AA992507	
C 91	17	77.3	254	9	AA911441	
C 92	17	77.3	254	9	AT327886	
C 93	17	77.3	254	13	BUS58101	
C 94	17	77.3	258	10	BG604516	
C 95	17	77.3	258	14	D76658	
C 96	17	77.3	282	10	BG606800	
C 97	17	77.3	285	10	BB598088	
C 98	17	77.3	286	10	BF46197	
C 99	17	77.3	286	14	CB361937	
C 100	17	77.3	286	28	BH229024	
C 101	17	77.3	269	28	BH229022	
C 102	17	77.3	273	28	AZ417494	
C 103	17	77.3	275	12	BJ284005	
C 104	17	77.3	279	13	BY120448	
C 105	17	77.3	283	10	BB855208	
C 106	17	77.3	283	14	D46518	
C 107	17	77.3	284	14	CA736089	
C 108	17	77.3	285	10	BE403060	
C 109	17	77.3	285	14	CA646584	
C 110	17	77.3	286	14	CA636766	
C 111	17	77.3	289	10	BF299427	
C 112	17	77.3	294	10	BE639324	
C 113	17	77.3	296	12	BI127904	
C 114	17	77.3	298	12	BI135859	
C 115	17	77.3	310	13	BY121347	
C 116	17	77.3	310	13	W42374	
C 117	17	77.3	313	28	AQ444327	
C 118	17	77.3	316	13	BY129327	
C 119	17	77.3	316	13	BY149919	
C 120	17	77.3	319	10	BG060942	

ALIGNMENTS

RESULT 1
BUS94332 900 bp mRNA linear EST 20-SEP-2002
LOCUS BUS94332
DEFINITION BUS94332.1 GI:23246091
ACCESSION BUS94332
VERSION BUS94332.1
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 900)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: gcaps-remail.nih.gov
Tissue Procurement: NCI
CDNA Library Preparation: Michael Brownstein Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov

FEATURES
source
Plate: L1CM2619 row: f column: 20
High quality sequence stop: 475.
Location/Qualifiers
1..900
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6449779"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (TI-Phase-resistant)"
/clone_id="NIH MGC 142"
/note="Vector: pDNR-LIB; Site 1: SfiI (ggccataggcc); Site 2: SfiI (ggccgctggcc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon - 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
5'-ATCTAGAGCCGACGCGCCGACATG-dT(30)NN-3'. Pull-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the >0.5 kb size fraction (other fractions present in NIH MGC 141). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH MGC library."

BASE COUNT 251 a 311 c 182 g 155 t 1 others
ORIGIN
Query Match 100.0%; Score 22; DB 13; Length 900;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTGGACACACGACGACCA 22
Db 298 ATTGGACACACGACGACCA 319

RESULT 2
CB447869 516 bp mRNA linear EST 26-MAR-2003
LOCUS CB447869
DEFINITION CB447869.1 GI:29254251
ACCESSION CB447869
VERSION CB447869.1
KEYWORDS
SOURCE
ORGANISM Bos taurus (cow)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 516)
AUTHORS Smith,T.P.L., Roberts,A.J., Scherrenkamp,S.E., Chitko-McKown,C.G., Wray,J.E. and Keeler,J.W.
TITLE A second set of bovine ESTs from pooled-tissue normalized libraries
JOURNAL Unpublished
COMMENT Contact: Smith TPJ
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smitht@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross_match v0.990329.
Plate: F0Y8056 row: D column: 21
Seq primer: TAGAAGCAGACGTCGAGG.

FEATURES
source
Location/Qualifiers
1..516
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"

```

/1ab_host="DH10B"
/clone_lib="MARC 6BOV"
/note="Vector: pCDNA3.1, Site 1: EcoRI, Site 2: NotI;
Library made with RNA pooled from multiple tissues
including liver, lung, hypothalamus, pituitary, and
placenta/endometrium."
BASE COUNT      91 a      114 c      171 g      140 t
ORIGIN
Query Match      90.9%; Score 20; DB 14; Length 516;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      3 TGGCACCACGACGACCA 22
|||||
Db      211 TGGCACCACGACGACCA 192

RESULT 3
LOCUS      CB447515      709 bp      mRNA      linear      EST 26-MAR-2003
DEFINITION      701522 MARC 6BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION      CB447515
VERSION      CB447515.1 GI:29253897
KEYWORDS      EST.
SOURCE      Bos taurus (cow)
ORGANISM      Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 709)
Smith,T.P.L., Roberts,A.J., Echeerikamp,S.E., Chitko-McKown,C.G.,
Wray,J.E. and Keeler,J.W.
A second set of bovine ESTs from pooled-tissue normalized libraries
Unpublished
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4350
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_al option. Vector identified with
cross_match v0.390329.
Plates: F018056 row: D column: 21
Seq primer: GTAATACGACTCACTATAGG.
Location/Qualifiers
1. 709
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/1ab_host="DH10B"
/clone_lib="MARC 6BOV"
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including liver, lung, hypothalamus, pituitary, and
placenta/endometrium."
BASE COUNT      189 a      220 c      157 g      143 t
ORIGIN
Query Match      90.9%; Score 20; DB 14; Length 709;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      3 TGGCACCACGACGACCA 22
|||||
Db      666 TGGCACCACGACGACCA 685

RESULT 4
LOCUS      B1675164/c      301 bp      mRNA      linear      EST 12-SEP-2001

```

```

DEFINITION      949079C01.X2 949 - Juvenile leaf and shoot cDNA from Steve Moose
Zea mays cDNA, mRNA sequence.
ACCESSION      B1675164
VERSION      B1675164.1 GI:15590548
KEYWORDS      EST.
SOURCE      Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 301)
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949079 row: C column: 01.
Location/Qualifiers
1. 301
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="W64A"
/db_xref="taxon:4577"
/tissue_type="Immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/1ab_host="E. coli XL0LR"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/note="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site 1: EcoRI; Site 2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled. PolyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridz vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia, 9 days after sowing. 4: Partially expanded and
greening leaves 4-5 at 13 days after sowing."
BASE COUNT      53 a      112 c      96 g      40 t
ORIGIN
Query Match      86.4%; Score 19; DB 12; Length 301;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      4 GCCACACGACGACGACCA 22
|||||
Db      271 GCCACACGACGACGACCA 253

RESULT 5
LOCUS      B1596214/c      370 bp      mRNA      linear      EST 07-SEP-2001
DEFINITION      949079C01.X1 949 - Juvenile leaf and shoot cDNA from Steve Moose
Zea mays cDNA, mRNA sequence.
ACCESSION      B1596214
VERSION      B1596214.1 GI:15499701
KEYWORDS      EST.
SOURCE      Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

```


REFERENCE 1 (bases 1 to 370)
 AUTHORS Walbot V
 TITLE Wize ESTs from various cDNA libraries sequenced at Stanford University
 JOURNAL Unpublished
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 723 8221
 Email: walbot@stanford.edu
 Plate: 949079 row: C column: 01.
 Location/Qualifiers
 source
 1..370
 /organism="Zea mays"
 /mol_type="mRNA"
 /cultivar="W64A"
 /db_xref="taxon:4577"
 /tissue_type="Immature leaf primordium and vegetative meristem"
 /dev_stage="4 stages from 3-13 days after imbibing"
 /lab_host="E. coli XL0R"
 /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve Moose"
 /note="Organ: juvenile vegetative shoots; Vector: pAD-GAL4-2.1; Site 1: EcoRI; Site 2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into lambda HybriTrap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing - includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia. 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."

BASE COUNT 82 a 129 c 111 g 48 t
 ORIGIN

Query Match 86.4%; Score 19; DB 12; Length 370;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GCCACCCACGACGACCA 22
 DB 340 GCCACCCACGACGACCA 322

RESULT 6
 A0158214/c 586 bp DNA linear GSS 12-SEP-1998
 LOCUS
 DEFINITION nbxb0010H08f CUGI Rice BAC Library Oryza sativa (japonica cultivar-group) genomic clone nbxb0010H08f, genomic survey sequence.

ACCESSION A0158214 GI:3592704
 VERSION A0158214
 KEYWORDS GSS.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 586)
 AUTHORS Wing, R.A. and Dean, R.A.
 TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
 JOURNAL Unpublished
 COMMENT On Sep 10, 1998 this sequence version replaced gi:355523.
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: TAAATGCACTCACTATAGG
 Class: BAC ends
 High quality sequence stop: 339.
 Location/Qualifiers
 source
 1..586
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="nbxb0010H08f"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"

/clone_lib="CUGI Rice BAC Library"
 /note="Vector: pReloBAC11; Site 1: HindIII; Site 2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 129 a 146 c 154 g 157 t
 ORIGIN

Query Match 86.4%; Score 19; DB 28; Length 586;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATTGCCACCCACGACGACCA 19
 DB 37 ATTGCCACCCACGACGACCA 19

RESULT 7
 A0290064/c 606 bp DNA linear GSS 03-DEC-1998
 LOCUS
 DEFINITION nbxb0036C08r CUGI Rice BAC Library Oryza sativa (japonica cultivar-group) genomic clone nbxb0036C08r, genomic survey sequence.

ACCESSION A0290064 GI:3951426
 VERSION A0290064.1
 KEYWORDS GSS.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 606)
 AUTHORS Wing, R.A. and Dean, R.A.
 TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
 JOURNAL Unpublished
 COMMENT On Sep 10, 1998 this sequence version replaced gi:355523.
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu

Seq primer: GGAAACAGCTATGACCATG
 Class: BAC ends
 High quality sequence stop: 447.
 Location/Qualifiers
 1. 606

FEATURES
 source
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="hdb0036C08r"
 /issue_type="leaf"
 /lab_host="E. coli DH10B"
 /clone_lib="CGI Rice BAC Library"
 /note="Vector: pBelBAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocytiledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT
 127 a 154 c 168 g 157 t

Query Match
 Best Local Similarity 86.4%; Score 19; DB 28; Length 606;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTGCACCAACGACGACA 19
 Db 38 ATTGCACCAACGACGACA 20

RESULT 8
 A0542963 627 bp DNA linear GSS 19-MAY-1999
 LOCUS RPTC1-11-368G5.TV RPTC1-11 Homo sapiens genomic clone RPTC1-11-368G5,
 DEFINITION genomic survey sequence.
 ACCESSION A0542963
 VERSION A0542963.1 GI:4873419
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and Venter
 'J.C.
 Use of BAC End Sequences from Library RPTC1-11 for Sequence-Ready
 Map Building
 Map Building
 Unpublished

JOURNAL
 COMMENT Other GSS: RPTC1-11-368G5.TV
 Contact: Shaying Zhao, William Niernan, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0208
 Fax: 301 838 0208
 Email: hbeew@igf.org

Clones are derived from the human BAC library RPTC1-11. For BAC
 library availability, please contact Pieter de Jong

(piet@delong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering/>) or from
 Research Genet cs (info@resgen.com). BAC end search page:
http://www.cigr.org/tcd/humgen/bac_end_search/bac_end_search.html.
 Seq primer: T7
 Class: BAC ends
 Location/Qualifiers
 1. 627

FEATURES
 source
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="GDB:7641076"
 /db_xref="taxon:9606"
 /clone="RPTC1-11-368G5"
 /sex="Male"
 /cell_type="lymphocytes"
 /clone_lib="RPTC1-11"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 RPTC11 Human Male BAC Library"
 BASE COUNT
 202 a 127 c 131 g 167 t

Query Match
 Best Local Similarity 86.4%; Score 19; DB 28; Length 627;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCAACGACGACCA 22
 Db 337 GCCACCAACGACGACCA 355

RESULT 9
 B1064968/c 629 bp mRNA linear EST 15-JUN-2001
 LOCUS pglfn.pk003.c15 normalized chicken fat cDNA library Gallus gallus
 DEFINITION cDNA clone pglfn.pk003.c15 5' similar to p1r/A30816/A30816 band 3
 anion transport protein (clone p1r1C1) - chickeng, mRNA sequence.
 ACCESSION B1064968
 VERSION B1064968.1 GI:14472490
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 629)
 COGburn,L.A., Morgan,R.W. and Burnside,J.
 Chicken ESTs from fat
 Unpublished

JOURNAL
 COMMENT Contact: Larry A. Cogburn
 University of Delaware
 Townsend Hall, Newark, DE 19717, USA
 Tel: 302-831-1335
 Fax: 302-831-2822
 Email: cogburn@udel.edu, www.chickest.udel.edu.
 Location/Qualifiers
 1. 629

FEATURES
 source
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /db_xref="taxon:9031"
 /clone="pglfn.pk003.c15"
 /sex="Male and Female"
 /tissue_type="fat"
 /lab_host="E.coli EMDH10B"
 /clone_lib="normalized chicken fat cDNA library"
 /note="Vector: pSPORT1"

BASE COUNT
 89 a 210 c 178 g 148 t 4 others

ORIGIN

Query Match
 Best Local Similarity 86.4%; Score 19; DB 12; Length 629;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 GCCACCAACGACGACCA 22

Db 372 GCCACACGACGACGACCA 354

RESULT 10

BF538979
LOCUS 60204852F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4187791 5', mRNA sequence.

ACCESSION

BF538979
VERSION BF538979.1 GI:11626360

KEYWORDS

EST.
Mus musculus (house mouse)

REFERENCE

1 (bases 1 to 7227)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished
Contact: Robert Strausberg, Ph.D.

Email: egads-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

CNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

FEATURES

source 1.1227

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"

/db_xref="taxon:10090"
/clone="IMAGE:4187791"

/lab_host="DH10B (T1 phage-resistant)"
/clone_1lb="NCI CGAP_SG2"

/note="Organ: salivary gland; Vector: PCMV-SPORE; Site: 1; NCI: Site 2; Salt: Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."

dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."

Technology. Note: this is a NCI CGAP Library."

BASE COUNT

304 a 393 c 361 g 168 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 77;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

3 TGGCACCACGACGACCA 21

Db

161 TGGCACCACGACGACCA 179

RESULT 11

AZ761910 34 bp DNA linear GSS 16-FEB-2001
LOCUS 1M0556C05R Mouse 10kb plasmid UGCM library Mus musculus genomic

DEFINITION

clone UGCM0556C05 R, genomic survey sequence.

ACCESSION

AZ761910
VERSION AZ761910.1 GI:12871328

KEYWORDS

GSS.
Mus musculus (house mouse)

ORGANISM

Mus musculus (house mouse)

REFERENCE

1 (bases 1 to 34)
Dunn, D., Aoyagi, A., Barber, M., Beacom, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weis, R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL

Unpublished
Contact: Robert B. Weiss
University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA

Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunne@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

Plate: 0556 row: C column: 05
Seq primer: CACACAGAAACACTATGAC

Class: plasmid ends
High quality sequence stop: 34.

FEATURES

source 1.34

/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"

/db_xref="taxon:10090"
/clone="UGCM0556C05"

/sex="male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/clone_1lb="Mouse 10kb plasmid UGCM library"
/note="Vector: PMD42uv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (G1473214|G1473214.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

and selected for ampicillin resistance."

TITLE
JOURNAL
COMMENT

Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
Development of Barley Transcriptome Resources
Unpublished
On Nov 21, 2001 this sequence version replaced gi:17027357.
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk
All sequence has a phred quality score of 20 or over
Seq primer: M13 reverse.
Location/Qualifiers

FEATURES
source

1..118
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="rRNA"
/cultivar="Optic"
/db_xref="taxon:112509"
/clone="Ebma07_S0001_K09"
/tissue_type="maternal"
/dev_stage="21 DPA"
/lab_host="DH10B"
/clone_lib="maternal", 21 DPA, no treatment, cv Optic,
EBma07"

BASE COUNT 25 a 52 c 35 g 6 t

ORIGIN

Query Match 81.8%; Score 18; DB 12; Length 118;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 CCACCACGACGACGACCA 22
|||||
21 CCACCACGACGACGACCA 38

RESULT 13
BI019327/c 207 bp mRNA linear EST 14-JUN-2001

LOCUS MR3-MT0328-080101-001-c03 MT0328 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI019327
VERSION BI019327.1 GI:14425957
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 207)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R.,
Nagai, M.A., da Silva, W. Jr., Zagro, M.A., Bordin, S., Costa, F.F.,
Golman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S.C., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.C., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.V. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202653
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922

Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAFESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3-MT0328-
080101-001-c03&ts=2001-01-08&tl=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 204.
Location/Qualifiers

FEATURES
source

1..207
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="MT0328"
/notes="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
'716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 20 a 75 c 62 g 50 t

ORIGIN

Query Match 81.8%; Score 18; DB 12; Length 207;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GCCACGACGACGACGAC 21
|||||
69 GCCACGACGACGACGAC 52

RESULT 14
BM376718 328 bp mRNA linear EST 23-JUL-2002

LOCUS EBem05_S0003_A14_R embryo, 14 DPA, no treatment, cv Optic, EBem05
DEFINITION Hordeum vulgare subsp. vulgare cDNA clone EBem05_S0003_A14 5', mRNA
sequence.
ACCESSION BM376718
VERSION BM376718.2 GI:21934299
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 328)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
Development of Barley Transcriptome Resources
Unpublished
On Jan 10, 2002 this sequence version replaced gi:18120108.
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk
All sequence has a phred quality score of 20 or over
Seq primer: M13 reverse.
Location/Qualifiers

FEATURES
source

1..328
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="rRNA"
/cultivar="Optic"
/db_xref="taxon:112509"
/clone="EBem05_S0003_A14"
/tissue_type="embryo"
/dev_stage="14 DPA"

```

/1ab host="DB10B"
/1ab lib="embryo, 14 DPA, no treatment, cv Optic,
EBen05"
/rtce="Vector: pSPORT1, Site_1: Sal I, Site_2: Not I,
Non-normalised library, directionally cloned into pSPORT1.
Derived from embryos dissected from developing grains (14
days post anthesis) in glasshouse grown barley plants.
Developed as part of the barley transcriptome resources of
BBSRC/SERAD funded cereal IGF (Investigating Gene
Function) project."
BASE COUNT      98 a      68 c      116 g      46 t
ORIGIN
Query Match      81.8%; Score 18; DB 12; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      5 CCACCACGACGACCA 22
|||||
179 CCACCACGACGACCA 196
Db
RESULT 15
BQ982962/c
LOCUS
DEFINITION
OGB17N11.yg abl OG_EFGH lettuce serriola Lactuca sativa cDNA clone
ACCESSION
BQ982962
VERSION
BQ982962.1 GI:22400487
KEYWORDS
EST.
SOURCE
Lactuca sativa
Lactuca sativa
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; eudicotyledons; Core eudicots;
asterids; campanulids; Asterales; Astereaceae; Cichoriaceae;
Cichorieae; Lactuca.
1 (bases 1 to 334)
Kozik,A., Van Dammé,M., Lavelle,D., Chevalier,P., Ziegler,U., Ellison
L.H., Van Dammé,M., Lavelle,D., Chevalier,P., Ziegler,U., Ellison
P., Koltman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
Church,S., Jackson,L. and Bradford,K
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Aamundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-11742
Fax: 1-(530)-752-9655
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig OG_Ca_Con15g2952, see http://cgdb.ucdavis.edu/
for details.
Plate: OGB17 row: N column: 11.
Location/Qualifiers
1..334
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="OGB17N11"
/lab_host="E.coli"
/clone_lib="OG_EFGH lettuce serriola"
/note="Vector: pBRCMDASf1ab. The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgdb.ucdavis.edu/
TAG LIB=OG_EFGH lettuce serriola

```

```

TAG: TISS=chemical induction
TAG_SEQ=TTGTAACCGGCG
BASE COUNT      91 a      62 c      95 g      86 t
ORIGIN
Query Match      81.8%; Score 18; DB 13; Length 334;
Best Local Similarity 100.0%; Pzed. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      5 CCACCACCAGACGACCA 22
      |||||
DB      86 CCACCACCAGACGACCA 69

RESULT 16      341 bp      mRNA      linear      EST 21-JUL-2000
LOCUS      BE398636
DEFINITION      WHE0023.D10P990702 TTBC WHE Wheat Endosperm library Triticum
ACCESSION      BE398636
VERSION      BE398636.1 GI:9358110
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum
Eukaryotes: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
1 (bases 1 to 341)
Triticum aestivum.
Triticum.
Triticaceae; Triticum.
REFERENCE      Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier,
AUTHORS      S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P.,
Herzmann, R.G., Holton, T., Jacquemais, J.M., Jia, J., Joudrier, P.,
Langridge, P., Lazo, G.R., Lin, J.Y., McGuire, P., Ogihara, Y.,
Pechioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M.,
Sorrells, M., Warburton, M., and Wenzel, G.
International Triticaceae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticaceae
Unpublished (2000)
JOURNAL
COMMENT      USDA ARS WRCR
      800 Buchanan Street, Albany, CA 94710-1105 USA
      Tel: 510 559 5773
      Fax: 510 559 5918
      Email: andersen@pw.usda.gov
      International Triticaceae EST Cooperative (ITEC)
      http://wheat.pw.usda.gov/genome.
      Location/Qualifiers
          1. 341
             /organism="Triticum aestivum"
             /mol_type="mRNA"
             /culivar="Cheyenne"
             /db_xref="taxon:4565"
             /clone_id="ITBC WHE Wheat Endosperm Library"
             /cissue_type="endosperm"
             /tissue_type="endosperm"
             /dev_stage="5-30 days post anthesis"
             /clone_id="ITBC WHE Wheat Endosperm Library"
             /note="Vector: Lambda ZAP1; Wheat Endosperm library
constructed in Lambda ZAP1 with 8-mer adapter."
          75 a      108 c      108 g      50 t

BASE COUNT
ORIGIN
Query Match      81.8%; Score 18; DB 10; Length 341;
Best Local Similarity 100.0%; Pzed. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      5 CCACCACCAGACGACCA 22
      |||||
DB      276 CCACCACCAGACGACCA 293

RESULT 17      386 bp      mRNA      linear      EST 06-DEC-2002
LOCUS      B0594281
DEFINITION      B012758-024-025-004-SP6 MP1Z-ADIS-024-developing root Beta vulgaris

```

ACCESSION BQ594281 CDNA clone 024-025-004 5-PRIME, mRNA sequence.
 VERSION BQ594281.1 GI:26123864
 KEYWORDS EST.
 SOURCE Beta vulgaris
 ORGANISM Beta vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Amaranthaceae; Beta.
 1 (bases 1 to 386)
 Herwig,R., Schulz,B., Weishaar,B., Hennig,S., Steinfath,M., Drungowski,M., Stahl,D., Wruck,W., Menze,A., O'Brien,J., Lehrach,H. and Radelof,U.
 Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes plant J. 32 (5), 845-857 (2002)
 TITLE ADIS DNA core facility at MPiZ
 JOURNAL Max-Planck-Institute for Plant Breeding Research
 COMMENT Carl-von-Linne Weg 10, 50829 Koeln, Germany
 Contact: Weishaar B
 Fax: 00492215062851
 Email: weishaar@mpi-z-koeln.mpg.de
 Insert Length: 386 Std Error: 0.00
 Plate: 25 Row: 0 Column: 04
 Seq primer: SP6; CATACGATTAGTGACACTATAG.
 Location/Qualifiers
 1..386
 /organism="Beta vulgaris"
 /mol_type="mRNA"
 /cultivar="KWS2320 (double haploid, monogerm breeding line)"
 /db_xref="GABI:192575"
 /db_xref="taxon:161934"
 /clone="024-025-004"
 /tissue_type="developing root"
 /lab_host="EMDH103"
 /clone_lib="MPiZ-ADIS-024-developing root"
 /note="Vector: PCWVS-PORT6; Site_1: SalI; Site_2: NotI; cDNA library from sugar beet, library provided by KWS Kleinzelleneber Saatnucht AG Einbeck, Germany, contact: b.schulz@kws.de; cloning sites SalI-NotI, primer sites and orientation:
 SP6-Sali-CCAGCGCCGCG-SPRIME-cDNA-polyA-CC-NotI-TT; Note: Sequencing granted in the context of the GABI-BEET project, local PI: Dr. Katharina Schneider, coordinator: Prof. Christian Jung; Sequence submission managed by RZPD/GABI-Primary database: <http://gabi.rzpd.de>"
 BASE COUNT 128 a 112 c 54 g 92 t
 ORIGIN

Query Match 81.8%; Score 18; DB 13; Length 386;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTGCCACGACGACG 18
 |||||
 61 ATTGCCACGACGACG 78
 |||||

RESULT 18
 LOCUS B1336904
 DEFINITION AR083075PAP075 Porcine Spleen cDNA library Sus scrofa cDNA, mRNA sequence.
 ACCESSION B1336904
 VERSION B1336904.1 GI:15418200
 KEYWORDS EST.
 SOURCE Sus scrofa (pigs)
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 389)
 Rink,A., Santschi,E.M. and Beatrice,C.W.

TITLE Amplified, Normalized cDNA Libraries from a Porcine Model of
 JOURNAL Orthopedic Implant Associated Staphylococcus aureus Infection
 COMMENT Unpublished
 Contact: Rink A
 Department of Animal Biotechnology
 College of Agriculture, Biotechnology and Natural Resources,
 University of Nevada, Reno
 MS 202, FA 105, 1664 N Virginia St, Reno, NV 89557-0236, USA
 Tel: 775 784 1375
 Fax: 775 784 1375
 Email: arink@cabmr.unr.edu
 Tissues and cells are derived from a porcine model for
 implant-associated infection using 1000 cfu of Staphylococcus
 aureus in a tibial transection, reduced and internally fixed with a
 dynamic compression plate. NOTE: The sequences contain a 'cDNA
 adapter' between the EcoRI site and the start of the EST. The
 adapter sequence is 'ATTCCGACGACG'.

FEATURES
 source
 1..389
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /strain="crossbreed"
 /db_xref="taxon:9823"
 /tissue_type="Spleen"
 /cell_type="mixed"
 /dev_stage="control, 5 month old castrated male"
 /lab_host="GOLR"
 /clone_lib="Porcine Spleen cDNA library"
 /note="Vector: pBSK; Site_1: Eco RI; Site_2: XhoI; Tissues and cells are derived from a porcine model for implant-associated infection using 1000 cfu of Staphylococcus aureus in a tibial transection, reduced and internally fixed with a dynamic compression plate. NOTE: The sequences contain a 'cDNA adapter' between the EcoRI site and the start of the EST. The adapter sequence is 'ATTCCGACGACG'."

BASE COUNT 120 a 118 c 68 g 72 t 11 others
 ORIGIN

Query Match 81.8%; Score 18; DB 12; Length 389;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACGACGACGACGCA 22
 |||||
 DB 215 CCACGACGACGACGCA 232
 |||||

RESULT 19
 LOCUS BQ593667
 DEFINITION E012763-024-A23-SP6 MPiZ-ADIS-024-developing root Beta vulgaris
 ACCESSION BQ593667
 VERSION BQ593667.1 GI:26123250
 KEYWORDS EST.
 SOURCE Beta vulgaris
 ORGANISM Beta vulgaris
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Amaranthaceae; Beta.
 1 (bases 1 to 396)
 Herwig,R., Schulz,B., Weishaar,B., Hennig,S., Steinfath,M., Drungowski,M., Stahl,D., Wruck,W., Menze,A., O'Brien,J., Lehrach,H. and Radelof,U.
 Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes plant J. 32 (5), 845-857 (2002)
 TITLE ADIS DNA core facility at MPiZ
 JOURNAL Max-Planck-Institute for Plant Breeding Research
 COMMENT Carl-von-Linne Weg 10, 50829 Koeln, Germany
 Contact: Weishaar B
 Fax: 00492215062851

FEATURES
source

Email: weischa@mpiz-koeln.mpg.de
Insert Length: 396 Std Error: 0.00
Plate: 26 Row: A Column: 23
Seq primer: SP6; CATACGATTAGTGACACTATAG.
Location/Qualifiers
1..396
/organism="Beta vulgaris"
/mol_type="rRNA"
/cultivar="KWS2320 (double haploid, monogerm breeding line)"
/db_xref="GABI:193194"
/db_xref="taxon:161934"
/clone="024-026-A23"
/issue_type="developing root"
/lab_host="EMDH10B"
/clone_lib="WPIZ-ADIS-024-developing root"
/note="Vector: PCWVSPORTE; Site 1: SalI; Site 2: NotI;
cDNA library from sugar beet, library provided by KWS
Kleinwanzlebener Saatgut AG Einbeck, Germany, contact:
b.schulz@kws.de; cloning sites SalI-NotI, primer sites and
orientation:
SP6-SalI-CCACGCGCTCCG-SPRIME-cDNA-polyA-CC-NotI-T7; Note:
Sequencing granted in the context of the GABI-Bset project
, local PI: Dr. Katharina Schneider, coordinator: Prof.
Christian Jung; Sequence submission managed by
RSPD/GABI-Primary database: <http://gabi.rzpd.de>"

BASE COUNT 128 a 132 c 53 g 82 t 1 others

Query Match 81.8%; Score 18; DB 13; Length 396;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 20 ATTGCCACCCAGCAGC 18
ATTGCCACCCAGCAGC 37

RESULT 20 399 bp mRNA linear EST 26-MAR-1999
C84299/c C84299 osteoclast substracted library Oryctolagus cuniculus cDNA.
LOCUS C84299 mRNA sequence.
DEFINITION C84299
ACCESSION C84299.1 GI:4527559
VERSION C84299.1
KEYWORDS EST.
SOURCE Oryctolagus cuniculus (rabbit)
ORGANISM Oryctolagus cuniculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
1 (bases 1 to 389)
Kobori, M., Ikeda, Y., Nara, H., Kato, M., Kamegawa, M., Nojima, H. and
Kawashima, H.
Large scale isolation of osteoclast-specific genes by an improved
method involving the preparation of a substracted cDNA library
Journal Genes Cells 3 (7), 459-475 (1998)
MEDLINE 98424349
PUBMED 9753427
COMMENT Contact: Kobori M
Molecular Medicine Laboratories
Institute for Drug Discovery Research, Yamanouchi Pharmaceutical
21, Miyukigaoka, Tsukuba, Ibaraki 305, Japan
Email: kobori@yamanouchi.co.jp
PROJECT = "OSG".
Location/Qualifiers
1..399
/organism="Oryctolagus cuniculus"
/mol_type="mRNA"
/db_xref="taxon:9986"
/tissue_type="long bone"
/cell_type="osteoclast"
/cell_line="primary"
/dev_stage="5 day-old"

BASE COUNT 74 a 104 c 144 g 68 t 9 others

Query Match 81.8%; Score 18; DB 14; Length 399;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 137 CCACCCAGCAGCAGCA 22
CCACCCAGCAGCAGCA 120

RESULT 21 406 bp DNA linear GSS 21-FEB-2001
A2859178/c 2M0164L07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
LOCUS clone UUGC2M0164L07 R. genomic survey sequence.
DEFINITION A2859178
ACCESSION A2859178.1 GI:13053221
VERSION A2859178.1
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 406)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, N., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
and Wright, D. Weis, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genome.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0164 Row: U Column: 07
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 406.
Location/Qualifiers
1..406
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0164L07"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: pMD42ny, Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into

FEATURES
source

Journal
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genome.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0164 Row: U Column: 07
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 406.
Location/Qualifiers
1..406
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0164L07"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: pMD42ny, Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into

Mon Aug 18 10:30:11 2003

us-10-074-620-6.c11.rst

Page 11

BASE COUNT 107 a 59 c 69 g 171 t
ORIGIN

Query Match 81.8%; Score 18; DB 28; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
5 CCACCACGACGACGACCA 22
396 CCACCACGACGACGACCA 379

RESULT 22
BI596215 410 bp mRNA linear EST 07-SEP-2001
LOCUS 949079C01.Y1 949 - Juvenile leaf and shoot cDNA from Steve Moose
DEFINITION Zea mays cDNA, mRNA sequence.
ACCESSION BI596215
VERSION BI596215.1 GI:15499702
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 410)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 8227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949079 row: C column: 01.
FEATURES
Source Location/Qualifiers
1..410
/organism="Zea mays"
/mol_type="mRNA"
/cultiVar="W64A"
/db_xref="taxon:4577"
/tissue_type="immature leaf primordium and vegetative meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve Moose"

notes="Organ: juvenile vegetative shoots; Vector: pAD-GAL4-2.1. Site 1: EcoRI; Site 2: XhoI; Equal amounts of total RNA by weight from 4 tissue sources (see below) were pooled, polyA+ RNA isolated, and cDNA synthesized for EcoRI (5') and XhoI (3') directional cloning into lambda HydriTap vector from Stratagene. Tissue Sources: 1. Whole shoots 3 days after sowing/imbibing in wet soil. 2. Basal 1.5 cm shoots 6 days after sowing includes yellow portions of developing leaves 1-5, primordia from 6-8, and the vegetative apex. 3. Non-green portions of developing leaves 4-5 and the vegetative apex, including adult leaf primordia, 9 days after sowing. 4. Partially expanded and greening leaves 4-5 at 13 days after sowing."
BASE COUNT 78 a 112 c 135 g 85 t
ORIGIN

Query Match 81.8%; Score 18; DB 13; Length 410;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
5 CCACCACGACGACGACCA 22

Db 9 CCACCACGACGACGACCA 26

RESULT 23
B0593650 417 bp mRNA linear EST 06-DEC-2002
LOCUS E012763-024-026-G23-SP6 MP12-ADIS-024-developing root Beta vulgaris
DEFINITION cDNA clone 024-026-G23 5-PRIME, mRNA sequence.
ACCESSION B0593650
VERSION B0593650.1 GI:26123233
KEYWORDS EST.
SOURCE Beta vulgaris
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllales; Amaranthaceae; Beta.
REFERENCE 1 (bases 1 to 417)
AUTHORS Herwig, R.; Schulz, B.; Weishaar, B.; Hennig, S.; Steinfach, M.; Drungowski, M.; Stahl, D.; Wreck, W.; Menze, A.; O'Brien, J.; Lehnach, H. and Radelof, U.
TITLE Construction of a 'unigene' cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes
JOURNAL Plant J. 32 (5), 845-857 (2002)
COMMENT Contact: Weishaar B
ADIS DNA core facility at MPIZ
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weishaar@mpiz-koeln.mpg.de
Insert Length: 417 Std Error: 0.00
Plate: 26 row: G column: 23
Seq primer: SP6; CATACGATTGAGTGACACTATAG.
FEATURES
Source Location/Qualifiers
1..417
/organism="Beta vulgaris"
/mol_type="mRNA"
/cultiVar="KWS2320 (double haploid, monogerm breeding line)"
/db_xref="GABI:193202"
/db_xref="taxon:161934"
/clone="024-026-G23"
/tissue_type="developing root"
/lab_host="EMDH10B"
/clone_lib="MPIZ-ADIS-024-developing root"
/note="Vector: PCWVS-PORT6; Site 1: SalI; Site 2: NotI; cDNA library from sugar beet, library provided by KWS Kleinfeldbecker Saat- und Zucht AG Einbeck, Germany; contact: b.schulz@kws.de; cloning sites SalI-NotI, primer sites and orientation:
SP6-SalI-CCACGCGCCG-5prime-cDNA-polyA-CC-NotI-T7. Note: Sequencing granted in the context of the GABI-beet project, local PI: Dr. Katharina Schneider, coordinator: Prof. Christian Jung; Sequence submission managed by RZPD/GABI-Primary database: http://gabi.rzpd.de"

BASE COUNT 134 a 140 c 54 g 86 t
ORIGIN

Query Match 81.8%; Score 18; DB 13; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATTGCGACGACGACGACG 18
21 ATTGCGACGACGACGACG 38
RESULT 24
CA723598 425 bp mRNA linear EST 26-NOV-2002
LOCUS wdr1f.pk003.j23 wdr1f Triticum aestivum cDNA clone wdr1f.pk003.j23
DEFINITION 5' end, mRNA sequence.
ACCESSION CA723598


```

VERSION          CA723598.1  GI:254455391
KEYWORDS
SOURCE
ORGANISM         Triticum aestivum (bread wheat)
                  Triticum aestivum
                  Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
REFERENCE
AUTHORS          1 (bases 1 to 425)
                  Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
                  Mao,G., Catharin,N. and Hanafey,W.K.
                  Dupont Wheat cDNA Sequence
TITLE            Unpublished
JOURNAL          Contact: Scott V. Tingey
COMMENT          Crop Genetics
                  E. I. Dupont de Nemours and Company
                  1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
                  Tel: 302-631-2602
                  Fax: 302-631-2607
                  Email: Scott.V.Tingey@USA.dupont.com
                  Seq primer: M13.
FEATURES
SOURCE
1..425
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wdrif.pk003.j23"
/tissue_type="root"
/lab_host="PH10B"
/clone_11b="wdrif"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Meat (Triticum aestivum)
length)"
BASE COUNT      79 a 144 c 135 g 59 t 8 others
ORIGIN
Query Match 81.8%; Score 18; DB 14; Length 425;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY 5 CCACCAACGAGCAGCACCA 22
    |||||
Db 24 CCACCAACGAGCAGCACCA 41
    |||||
RESULT 25
LOCUS           A0042664 436 bp DNA linear GSS 14-JUL-1998
DEFINITION      CIT-HSP-2326K17.TR CIT-HSP Homo sapiens genomic clone 2326K17,
                genomic survey sequence.
ACCESSION       A0042664
VERSION         A0042664.1 GI:2310049
KEYWORDS        GSS.
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
                1 (bases 1 to 436)
                Adams,M.D., Rounaley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
                Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
                Simon,M. and Venter,J.C.
                Use of a random BAC End Sequence Database for Sequence-Ready Map
                Building (1998)
                Unpublished
                Cite_GSSs: CIT-HSP-2326K17.TV
                Contact: Mark Adams
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: mdadams@igf.org
                Clones are available from Research Genetics (info@resgen.com). BAC
                end search page:

```

```

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends
FEATURES             Location/Qualifiers
     source            1..436
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /clone="2326K17"
                        /sex="Male"
                        /cell_type="Sperm"
                        /clone_1fb="CIT-HSP"
                        /note="Vector: pbelBAC11, Site_1: HindIII, Site_2: HindIII"
BASE COUNT           105 a      112 c      67 g      150 t      2 others
ORIGIN
Query Match          81.8%; Score 18; DB 28; Length 436;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy                   3 TGGCCACCACGAGCAGCAC 20
Db                  196 TGCACCAACGACGACGAC 213
RESULT 26
LOCUS                BQ982501               437 bp    mRNA       linear   EST 21-AUG-2002
DEFINITION            QG16U21.yg.abi QG_EFGHU lettuce serritola Lactuca sativa cDNA clone
VERSION               QG16U21, mRNA sequence.
KEYWORDS              BQ982501
SOURCE                BQ982501.1 GI:22400026
ORGANISM              Est.
                     Lactuca sativa
                     Lactuca sativa
                     Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                     Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                     asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
                     Cichorieae; Lactuca.
REFERENCE              1 (bases 1 to 437)
AUTHORS               Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
                     Lin,H., van Damme-More, Lavelle,D., Chevallier,P., Ziegler,J., Ellison
                     ,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
                     Church,S., Jackson,L. and Bradford,K.
COMMENT               Lettuce and Sunflower ESTs from the Compositeae Genome Project
                     http://compgenome.ucdavis.edu/
                     Unpublished
                     Contact: Alexander Kozik [R.W.Michelmore]
                     Department of Vegetable Crops, R.W.Michelmore Lab
                     University of California at Davis (UCD)
                     Asmundson Hall, UCD, Davis, CA 95616, USA
                     Tel: 1-(530)-742-1742
                     Fax: 1-(530)-752-3655
                     Email: akozik@ucdavis.edu [michelmoreveg@mail.ucdavis.edu]
                     singleton, see http://cgpdb.ucdavis.edu/ for details.
                     Plate: QG616 row: J column: 21.
FEATURES             Location/Qualifiers
     source            1..437
                        /organism="Lactuca sativa"
                        /mol_type="RNA"
                        /culivar="L.serritola"
                        /DB_xref="taxon:4236"
                        /clone="QG16U21"
                        /lab_host="E.coli"
                        /clone_1fb="QG_EFGHU lettuce serritola"
                        /note="Vector: pBRDNASfiIAB; The library was constructed
                        from 10 different sources of RNA from a single genotype.
                        Separate cDNAs were generated using primers that
                        incorporated unique 5' and 3' tags to distinguish each
                        source of RNA. cDNAs were then pooled, size-fractionated,
                        directionally cloned into a custom medium-copy vector and
                        transformation made with four size classes to minimize

```

size bias. Details of each source of RNA and library construction can be obtained at http://cspdb.ucdavis.edu/TAG_LIB-OG_EFGH lettuce serricola
TAG_TISSUE=callus
TAG_SEQ=GTACTGCCGCG

BASE COUNT 109 a 45 c 153 g 129 t 1 others
ORIGIN

Query Match 81.8%; Score 18; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 CCACCACGACGACCA 22
Db 189 CCACCACGACGACCA 172

RESULT 27
AZ234554 443 bp DNA linear GSS 14-JUN-2000
LOCUS RPCI-23-55L21.TU RPCI-23 Mus musculus genomic clone RPCI-23-55L21,
DEFINITION genomic survey sequence.
ACCESSION AZ234554
VERSION AZ234554.1 GI:8542600
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 443)
Zhuo, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Akhmet
and Fraser, C.W.
B., Levine, M., Megann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
'B', Levine, M., Megann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
Mouse BAC End Sequences from library RPCI-23
Unpublished
Other_GSSs: RPCI-23-55L21.TU
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

TITLE
JOURNAL
COMMENT
Other_GSSs: RPCI-23-55L21.TU
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 55 row: L column: 21
Seq primer: SP6
Classes: BAC ends.
Location/Qualifiers

1. 443
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-55L21"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"

BASE COUNT 104 a 81 c 101 g 157 t
ORIGIN
Query Match 81.8%; Score 18; DB 28; Length 443;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 CCACCACGACGACCA 22
Db 422 CCACCACGACGACCA 405

RESULT 28
B0593660 447 bp mRNA linear EST 06-DEC-2002
LOCUS E012763-024-026-E23-SP6 MP12-ADIS-024-developing root Beta vulgaris
DEFINITION cDNA clone 024-026-E23 5-PRIME, mRNA sequence.
ACCESSION B0593660
VERSION B0593660.1 GI:26123243
KEYWORDS EST.
SOURCE Beta vulgaris
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllales; Amaranthaceae; Beta.
1 (bases 1 to 447)
Herwig, R., Schulz, B., Weisshaar, B., Hennig, S., Steinfath, M.,
Drungowski, M., Stahl, D., Wruick, W., Wenzel, A., O'Brien, J., Lehrach, H.
and Radeke, U.
Construction of a 'unigene' cDNA clone set by oligonucleotide
fingerprinting allows access to 25 000 potential sugar beet genes
Plant J. 32 (5), 845-857 (2002)
Contact: Weisshaar B
ADIS DNA core facility at MP12
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weisshaar@mpiz-koeln.mpg.de
Insert Length: 447 Std Error: 0.00
Plate: 26 row: E column: 23
Seq primer: SP6, CATGAGTTTGGTCACTATG.
Location/Qualifiers

TITLE
JOURNAL
COMMENT
ADIS DNA core facility at MP12
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weisshaar@mpiz-koeln.mpg.de
Insert Length: 447 Std Error: 0.00
Plate: 26 row: E column: 23
Seq primer: SP6, CATGAGTTTGGTCACTATG.
Location/Qualifiers

FEATURES
source
1. 447
/organism="Beta vulgaris"
/mol_type="mRNA"
/cultivar="KWS2320 (double haploid, monogerm breeding line)"
/db_xref="GABI:193199"
/db_xref="taxon:161934"
/clone="024-026-E23"
/tissue_type="developing root"
/lab_host="EMDH10B"
/clone_lib="MP12-ADIS-024-developing root"
/note="Vector: PCWVS-POR6; Site 1: SalI; Site 2: NotI;
cDNA library from sugar beet library provided by KWS
Kleinwanzlebener Saatgut AG Einbeck, Germany, contact:
b.schulz@kws.de; Cloning sites SalI-NotI, primer sites and
orientation:
SP6-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-17; Note:
Sequencing granted in the context of the GABI-Best project
local PI: Dr. Katharina Schneider, coordinator: Prof.
Christian Jung; Sequence submission managed by
RZPD/GABI-Primary database: <http://gabi.rzpd.de>"
RZPD/GABI-Primary database: <http://gabi.rzpd.de>"

BASE COUNT 147 a 144 c 59 g 96 t
ORIGIN

Query Match 81.8%; Score 18; DB 13; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATTGCCACGACGACGACG 18
Db 60 ATTGCCACGACGACGACG 77

RESULT 29
B0395258/c

Mon Aug 18 10:30:11 2003

us-10-074-620-6.011.rst

Page 14

LOCUS BQ395258 474 bp mRNA linear EST 22-MAY-2002
DEFINITION NISC ngl14906.y1 NICHD_XCC_Emb6 Silurana tropicalis cDNA clone
IMAGE:3383187.5', mRNA sequence.
ACCESSION BQ395258
VERSION BQ395258.1 GI:21082945
KEYWORDS EST.
SOURCE Silurana tropicalis (western clawed frog)
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Silurana.
REFERENCE
AUTHORS NIH-XCC http://image.llnl.gov/image/html/xenopuslib info.shtml.
1 (bases 1 to 474)
TITLE National Institute of Child Health and Human Development, National
Cancer Institute, Xenopus Gene Collection
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
CDNA Library Preparation:
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
info@image.llnl.gov
Plate: LLM11976 row: A column: 12
Seq primer: M13P1 reverse primer (ABI).
Location/Qualifiers
1..474
/organism="Silurana tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:5383187"
/tissue_type="neutula"
/dev_stage="embryo, stages 14-19"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD XCC Emb6"
/note="Vector: pCMV-SF0RT6.ccd, Site_1: NotI, Site_2:
EcoRV; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 2.1 kb. Constructed by Invitrogen. Note: This
is a Xenopus Gene Collection (XGC) library."

BASE COUNT 137 a 108 c 118 g 111 t
ORIGIN

Query Match 81.8%; Score 18; DB 13; Length 474;
Best Local Similarity 100.0%; Freq. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACCACGACGACCA 22
DB 405 CCACCACGACGACCA 388

RESULT 30
LOCUS BQ796727 495 bp mRNA linear EST 30-JUL-2002
DEFINITION EST 5665 Ripening Grape berries lambda zap II library Vitis
vitis cDNA clone RT043610.3', mRNA sequence.
ACCESSION BQ796727
VERSION BQ796727.1 GI:22011693
KEYWORDS EST.
SOURCE Vitis vinifera
ORGANISM Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; Vitaceae; Vitis.
REFERENCE
AUTHORS Abbal, P., Agasse, A., Ageorges, A., Atanasova, R., Barrieu, F.,
Couture, C., Dedaldecamp, F., Delrot, S., Glissant, D., Grimplet, J.,
Hamdi, S., Romieu, C. and Terrier, N.
TITLE Generation of Expressed Sequence Tag from Grape Berry (skin, pulp
or seeds) at Various Developmental Stages
JOURNAL Unpublished

COMMENT Contact: Romieu C.
Unite de Recherche des Produits de la Vigne
Institut National de la Recherche Agronomique
2, place Viala, 34 060 Montpellier Cedex 01, France
Tel: 00-33-(0)4-99-61-28-62
Fax: 00-33-(0)4-99-61-28-57
Email: romieu@enscm.inra.fr
Seq primer: 7'
Location/Qualifiers
1..495
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Shiraz"
/db_xref="taxon:29760"
/clone="RT043610"
/dev_stage="ripening stage"
/clone_lib="Ripening Grape berries lambda zap II library"
/note="Organ: Fruit; Vector: Lambda Zap II; Site 1: Eco RI
; Site 2: XhoI; Oriented library, construction described
in Generation of ESTs from grape berry (skin, pulp or
seeds) at various developmental stages by Terrier, N.,
Ageorges, A., Abbal, P., Romieu, C. in J. Plant Physiol. 158
(12): 1575-83 2001"

BASE COUNT 134 a 112 c 108 g 141 t
ORIGIN

Query Match 81.8%; Score 18; DB 13; Length 495;
Best Local Similarity 100.0%; Freq. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACCACGACGACCA 22
DB 456 CCACCACGACGACCA 473

RESULT 31
LOCUS CNS07HT1 498 bp DNA linear GSS 03-OCT-2001
DEFINITION Anopheles gambiae GSS SP6 end of clone 30109 of library NotreDame1
from strain PST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL611479.1 GI:15962902
VERSION GSS.
KEYWORDS Anopheles gambiae (African malaria mosquito)
SOURCE Anopheles gambiae
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 498)
Genoscope.
Direct Submission
Submitted (01-OCT-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
2 (bases 1 to 498)
Roth, C.W., Brey, P.T., Ke, Z. and Collins, F.H.
Direct Submission
Submitted (01-OCT-2001) BWHI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
Location/Qualifiers
1..498
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PST"
/db_xref="taxon:7165"
/clone="30109"
/clone_lib="NotreDame1"
/note="End : SP6"

FEATURES
source

Mon Aug 18 10:30:11 2003

us-10-074-620-6.011.rst

Page 15

BASE COUNT 119 a 133 c 117 g 128 t 1 others
ORIGIN

Query Match 81.8%; Score 18; DB 29; Length 498;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACCACGACGACGACCA 22
DB 121 CCACCACGACGACGACCA 138

RESULT 32
BH006015/c 521 bp DNA linear GSS 04-MAY-2001

DEFINITION BMBAC10J0777 PSU Brugia malayi Genomic Bac Library 1 & 2 Brugia
malayi genomic, genomic survey sequence.

ACCESSION BH006015
KEYWORDS GSS: +44 13948084

SOURCE Brugia malayi
ORGANISM Brugia malayi
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.

REFERENCE 1 (bases 1 to 521)
Whitton,C., Dago,J., Mare,J., Quail,M., Hall,N., Barrell,B., Foster
,J., Guillano,D., Slatko,B. and Blaxter,M.

TITLE Genome survey sequences from the human parasitic nematode Brugia
malayi

JOURNAL Unpublished
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk

Sequenced from the Filarial Genome Project's Brugia malayi BAC
library constructed by Jesse Pope-Chappel and Jeremy Foster. The
sequence was generated by The Pathogen Sequencing Unit, The Sanger
Centre, Cambridge, UK in collaboration with Mark Blaxter, ICAFP
University of Edinburgh, Edinburgh, UK
Seq primer: T7 (TAATACGACTCACTATAGG)
Class: BAC ends.

FEATURES
source Location/Qualifiers

1..521
/organism="Brugia malayi"
/mol_type="genomic DNA"
/strain="T8"
/db_xref="taxon:5279"
/sex="Mixed (male and female)"
/tissue_type="whole parasite"
/dev_stage="adult"
/clone_lib="Brugia malayi Genomic Bac Library 1 & 2"
/note="Vector: pSelecBAC II; Site 1: Hind III; Brugia
malayi genomic DNA was partially cleaved with Hind III and
size fractionated. 18,000 clones were generated from 2
libraries with mean insert size 60 kbp. The library was
constructed by Jesse Pope-Chappel, Smith College
Northampton MA and Dr Jeremy Foster, New England Biolabs,
MA."

BASE COUNT 129 a 78 c 129 g 185 t
ORIGIN

Query Match 81.8%; Score 18; DB 28; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCACCACGACGACGACCA 22
DB 516 CCACCACGACGACGACCA 499

RESULT 33
BQ593655 527 bp mRNA linear EST 06-DEC-2002

DEFINITION B012763-024-026-K23-SP6 MP12-ADIS-024-developing root Beta vulgaris
cDNA clone 024-026-K23 5-PRIME, mRNA sequence.

ACCESSION BQ593655
VERSION BQ593655.1 GI:26123238
KEYWORDS EST.

SOURCE Beta vulgaris
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllales; Amaranthaceae; Beta.

REFERENCE 1 (bases 1 to 527)
Herwig,R., Schulz,B., Weissnar,B., Hennig,S., Steinfath,M.,
Drungowski,M., Stahl,D., Wruick,W., Menze,A., O'Brien,J., Lebrach,H.
and Radloff,U.

TITLE Construction of a 'unigene' cDNA clone set by oligonucleotide
fingerprinting allows access to 25 000 potential sugar beet genes
Plant J. 32 (5), 845-857 (2002)

CONTACT: Weisshar B
ADIS DNA core facility at MP12
Max-Planck-Institute for Plant Breeding Research
Carl-von-Linne Weg 10, 50829 Koeln, Germany
Fax: 00492215062851
Email: weissha@mp12-koeln.mpg.de

Insert length: 527 Std Error: 0.00
Plate: 26 row: K column: 23
Seq primer: SP6; CATACGATTTAGCTGACACTATAG.

FEATURES
source Location/Qualifiers

1..527
/organism="Beta vulgaris"
/mol_type="mRNA"
/cultivar="KMS2320 (double haploid, monogerm breeding line
)"
/db_xref="GABI:193207"
/db_xref="taxon:161934"
/clone="024-026-K23"
/tissue_type="developing root"
/lab_host="EMDH10B"
/clone_lib="MP12-ADIS-024-developing root"
/note="Vector: pCMVSPORT6; Site 1: SalI; Site 2: NotI;
cDNA library from sugar beet, library provided by KMS
Kleinwanzlebener Saatgut AG Einbeck, Germany, contact:
b.schulz@kws.de; cloning sites SalI-NotI, primer sites and
orientation:
SP6-SalI-CCACGCGTCGCG-5prime-cDNA-polyA-Cc-NotI-T7; Note:
Sequencing granted in the context of the GABI-Best project
, local PI: Dr. Katharina Schneider, coordinator: Prof.
Christian Jung; Sequence submission managed by
RZPD/GABI-Primary database: http://gabi.rzpd.de"

BASE COUNT 172 a 164 c 76 g 115 t
ORIGIN

Query Match 81.8%; Score 18; DB 13; Length 527;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTGCCACGACGACGACG 18
DB 60 ATTGCCACGACGACGACG 77

RESULT 34

BQ282622 529 bp mRNA linear EST 13-MAY-2002

DEFINITION WHE3074_F06_L122S wheat cold-stressed seedling subcloned cDNA
library Triticum aestivum cDNA clone WHE3074_F06_L12, mRNA
sequence.

ACCESSION BQ282622
VERSION BQ282622.1 GI:20551304
KEYWORDS EST.

Mon Aug 18 10:30:11 2003

us-10-074-620-6.011.rst

Page 16

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
Location/Qualifiers
1. 529
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WAB3074_P06_L12"
/tissue_type="Seedling"
/dev_stage="Five-day Old seedling"
/lab_host="E. coli SOLR"
/clone_1lb="Wheat cold-stressed seedling subtracted cDNA library"
/note="Vector: Lambda Uni-Zap XR, excised phagemid, Site 1: EORI, Site 2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water/mystatin and cefotaxime in covered crystallization dishes. Five-day old seedlings were transferred to 5 C cold room and kept for 48 hr. The tissue, total RNA, and poly(A) RNA were prepared. A cDNA library was made, and the cDNA clones were in vivo excised to give pluscript phagemids in the 10 Close lab (Choi, Close, Fenton) at the University of California, Riverside. The cDNA clones were in vivo excised to give pluscript phagemids before subtraction was carried out. The mass excision of phagemid library and subtraction were done in HT Nguyen lab by D. Zhang at Texas Tech University. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT
ORIGIN
101 a 175 c 184 g 69 t
Query Match
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 GCCACCCACGACGACGAC 21
DB 178 GCCACCCACGACGACGAC 195
RESULT 35
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BOS93645
E012763.024-026-M23-SP6 MP12-ADIS-024-developing root Beta vulgaris
CDNA clone 024-026-M23 5-PRIME, mRNA sequence.
BOS93645
BOS93645.1 GI:26123228
EST.
Beta vulgaris
Beta vulgaris

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Amaranthaceae; Beta.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
Location/Qualifiers
1. 534
/organism="Beta vulgaris"
/mol_type="mRNA"
/cultivar="KWS2320 (double haploid, monogerm breeding line)"
/db_xref="GABI:193210"
/db_xref="taxon:161934"
/clone="024-026-M23"
/tissue_type="developing root"
/lab_host="EMDH10B"
/clone_1lb="MP12-ADIS-024-developing root"
/note="Vector: PCMVSPORT6; Site 1: SalI; Site 2: NotI; cDNA library from sugar beet library provided by KWS Kleinfanzlebeener Saatzeuche AG Birkbeck, Germany, contact: b.schulz@kws.de; cloning sites SalI-NotI, primer sites and orientation: SP6-SalI-CCACGCGTCCG-5prime-cDNA-polyA-CC-NotI-T7; Note: Sequencing granted in the context of the GABI-Beet project, local PI: Dr. Katharina Schneider, coordinator; Prof. Christian Jung; Sequence submission managed by RZPD/GABI-Primary database: <http://gabi.rzpd.de>"
BASE COUNT
ORIGIN
175 a 165 c 81 g 113 t
Query Match
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATTGCCACCCACGACGAC 18
DB 60 ATTGCCACCCACGACGAC 77
Search completed: August 15, 2003, 10:58:06
Job time : 1386.2 secs

Mon Aug 18 10:30:10 2003

us-10-074-620-6.01i.rge

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:32:41 ; Search time 602.525 Seconds

(without alignments)
1493.734 Million cell updates/sec

Title: US-10-074-620-6

Sequence: 1 attgccaccaccagcaccac 22

Scoring table: OLIGO_NUC

Searched: 2888711 seqs, 20454813386 residues

Word size

Total number of hits satisfying chosen parameters: 5777422

```
Minimum DB seq length: 0
```

Post-processing: Listing first 120 summaries

Database :

1: gb_ba:*
 2: gb_hng:*
 3: gb_in:*
 4: gb_cm:*
 5: gb_ov:*
 6: gb_pat:*
 7: gb_ph:*
 8: gb_pl:*
 9: gb_pr:*
 10: gb_ro:*
 11: gb_sts:*
 12: gb_sy:*
 13: gb_un:*
 14: gb_vl:*
 15: em_ba:*
 16: em_fua:*
 17: em_hum:*
 18: em_in:*
 19: em_mu:*
 20: em_cm:*
 21: em_ov:*
 22: em_ov:*
 23: em_pat:*
 24: em_pl:*
 25: em_pl:*
 26: em_ro:*
 27: em_ste:*
 28: em_un:*
 29: em_un:*
 30: em_hng_hum:*
 31: em_hng_inv:*
 32: em_hng_other:*
 33: em_hng_mus:*
 34: em_hng_pin:*
 35: em_hng_rod:*
 36: em_hng_mam:*
 37: em_hng_vrt:*
 38: em_hng_vrt:*
 39: em_hng_hum:*
 40: em_hng_mus:*
 41: em_hng_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	22	100.0	22	6	AX522241	AX522241 Sequence
2	22	100.0	5050	14	HS402IRA	K03333 Epstein-Barr
3	22	100.0	6011	14	HS4RSIR2T	J02070 Epstein-Barr
4	22	100.0	171823	14	HHV507799	AJ507799 Human hec
5	22	100.0	172281	14	EBV	V01555 Epstein-Barr
6	22	100.0	184113	14	HS4BS58BAJ	M80517 Epstein-Barr
7	20	90.9	266	14	S71027	S71027 nuclear pcr
8	20	90.9	650	11	BV031762	BV031762 src
9	20	90.9	14328	5	AC009695	AC009695 Homo sapi
10	20	86.4	1272	5	AF247362	AF247362 Petromyzo
11	19	86.4	2001	6	AX078616	AX078616 Sequence
12	19	86.4	2028	6	AX078615	AX078615 Sequence
13	19	86.4	2847	5	AF109781	AF109781 Dario rer
14	19	86.4	2981	1	AF333961	AF333961 Pseudomon
15	19	86.4	3007	5	CHKATE	M33404 Chicken ery
16	19	86.4	3407	5	CHKATE	M19496 Chicken ery
17	19	86.4	9359	2	AC018248	AC018248 Drosophi
18	19	86.4	26545	9	AC142283	AC142283 Homo sapi
19	19	86.4	75022	9	AF152638	AF152638 Homo sapi
20	19	86.4	110000	2	LMFLCHR33_02	Contribution (3) of
21	19	86.4	134582	8	AP003379	AP003379 Oryza sat
22	19	86.4	146101	9	AC009720	AC009720 Homo sapi
23	19	86.4	146670	8	AP003350	AP003350 Oryza sat
24	19	86.4	151703	2	AP003442	AP003442 Oryza sat
25	19	86.4	156964	9	AL450333	AL450333 Human DNA
26	19	86.4	163990	2	AC012132	AC012132 Homo sapi
27	19	86.4	164814	9	AC027801	AC027801 Homo sapi
28	19	86.4	172869	3	AC012160	AC012160 Drosophi
29	19	86.4	172860	9	AC136968	AC136968 Pan trogl
30	19	86.4	176258	2	AC126925	AC126925 Canis fam
31	19	86.4	177793	3	AC012096	AC012096 Drosophi
32	19	86.4	181062	2	AC062006	AC062006 Homo sapi
33	19	86.4	215050	1	AL646084	AL646084 Ralstonia
34	19	86.4	234920	2	AC126592	AC126592 Rattus no
35	19	86.4	245620	2	AC099575	AC099575 Rattus no
36	19	86.4	250030	2	AC129281	AC129281 Rattus no
37	19	86.4	257837	2	AC096594	AC096594 Rattus no
38	19	86.4	258439	2	AC094968	AC094968 Rattus no
39	19	86.4	298641	1	AB003504	AB003504 Drosophi
40	19	86.4	300099	1	AB016764	AB016764 Escherich
41	19	86.4	305245	1	AB016793	AB016793 Pseudomon
42	18	81.8	123	6	AX088696	AX088696 Sequence
43	18	81.8	227	5	AY022665	AY022665 Oryza sat
44	18	81.8	362	5	AF455324	AF455324 Hemidagru
45	18	81.8	398	8	AF317659	AF317659 Candida a
46	18	81.8	951	3	AF144262	AF144262 Caenorhab
47	18	81.8	1158	6	AX193986	AX193986 Sequence
48	18	81.8	1167	6	AX193992	AX193992 Sequence
49	18	81.8	1167	6	AX193996	AX193996 Sequence
50	18	81.8	1302	1	AX254170	AX254170 Pseudomon
51	18	81.8	1518	6	AX088675	AX088675 Sequence
52	18	81.8	1518	6	AX451734	AX451734 Sequence
53	18	81.8	1518	12	SC0313179	AJ333179 Synthetic
54	18	81.8	1668	6	AX562261	AX562261 Sequence
55	18	81.8	1842	3	AF144257	AF144257 Caenorhab
56	18	81.8	2002	6	AX594570	AX594570 Sequence
57	18	81.8	2110	5	AB052339	AB052339 Lampetra
58	18	81.8	2200	6	AX451740	AX451740 Sequence
59	18	81.8	2562	5	AF144425	AF144425 Gallus ga
60	18	81.8	2562	5	AF246959	AF246959 Gallus ga
61	18	81.8	2936	5	AF201076	AF201076 Gallus ga
62	18	81.8	3560	5	AF132531	AF132531 Gallus ga
63	18	81.8	3581	5	AB025831	AB025831 Coccini
64	18	81.8	4100	6	BD125565	BD125565 Polynucle
65	18	81.8	4100	14	SHUDNNG	L18355 Pseudorabie

ALIGNMENTS

ALIGNMENTS

RESULT 1				
LOCUS	AX522241			
DEFINITION	AX522241	22 bp	DNA	linear
ACCESSION	AX522241	Sequence 6 from Patent WO02064842.		PAT 24-OCT-2002
VERSION	AX522241			
KEYWORDS	AX522241.1	GI 72411119		
SOURCE				
ORGANISM		Human herpesvirus 4 (Epstein-Barr virus)		
		Human herpesvirus 4		
		Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
		Gammaherpesvirinae; Lymphocryptovirus.		
REFERENCE				
AUTHORS				
		Witte, D. P. and Groen, P. A.		

[illegible]

GLHGGGAGDSPTPGSNAPVCRNSHTATPNVSPHIEPESHSNPEAPILFPDDWYPP
SIDPADLDESDMYIFETESPSPSDDYEGHSKPRPSIQ"

repeat_region 3581..3618
/note="IR2 repeat partial copy"

repeat_region 3619..3743
/note="IR2 repeat copy A"

repeat_region 3744..3868
/note="IR2 repeat copy B"

repeat_region 3869..3993
/note="IR2 repeat copy C"

repeat_region 3994..4118
/note="IR2 repeat copy D"

repeat_region 4119..4243
/note="IR2 repeat copy E"

repeat_region 4244..4368
/note="IR2 repeat copy F"

repeat_region 4369..4493
/note="IR2 repeat copy G"

repeat_region 4494..4618
/note="IR2 repeat copy H"

repeat_region 4619..4743
/note="IR2 repeat copy I"

repeat_region 4744..4868
/note="IR2 repeat copy J"

repeat_region 4869..4993
/note="IR2 repeat copy K"

BASE COUNT 851 a 1830 c 1417 g 952 t

ORIGIN 1 bp upstream of BamHI site.

Query Match 100.0%; Score 22; DB 14; Length 5050;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTGCCACCACGACGACCA 22
Db 2368 ATTGCCACCACGACGACCA 2409

RESULT 3
HS4RSIRZ2T 6011 bp DNA linear VRL 02-AUG-1993
DEFINITION Epstein-Barr virus (wild type B95-8) DL and DR segments containing
IR2 repeats.
ACCESSION J02070.1 GI:330432
VERSION J02070
KEYWORDS Repeat region; tandem repeat.
SOURCE Human herpesvirus 4 (Epstein-Barr virus)
ORGANISM Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1894 to 2018)
Dambaugh, T.R. and Kieff, E.
Identification and nucleotide sequences of two similar tandem
direct repeats in Epstein-Barr virus DNA
J. Virol. 44 (3), 823-833 (1982)

JOURNAL MEDLINE 83085966
PUBMED 6294332
REFERENCE 2 (bases 1533 to 2018; 2922 to 4098)
AUTHORS Jeang, K.T. and Hayward, S.D.
TITLE Organization of the Epstein-Barr virus DNA molecule. III. Location
of the P1HR-1 deletion junction and characterization of the NcRI
repeat units that form part of the template for an abundant
12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
J. Virol. 48 (1), 135-148 (1983)

JOURNAL MEDLINE 83294686
PUBMED 6310141
REFERENCE 3 (bases 1533 to 2018; 2922 to 4098)
AUTHORS Jeang, K.-T. and Hayward, S.D.
TITLE Unpublished (1984)
REFERENCE 4 (bases 1 to 6011)
AUTHORS Jones, M.D., Foster, L., Sheedy, T. and Griffin, B.E.
The EB virus genome in Daudi Burkitt's lymphoma cells has a
deletion similar to that observed in a non-transforming strain

(P3HR-1) of the virus
EMBO J 3 (4), 813-821 (1984)

JOURNAL MEDLINE 84207939
PUBMED 6327290
COMMENT Original source text: Epstein-Barr virus (strain B95-8) DNA.
There are 12 and 1/3 (125 bp) tandem repeats found in the BamHI-H
fragment of the EBV B95-8 genome. It is thought that they play a
regulatory role. A 2.5 kb open reading frame overlapping the
repeats is also found in this fragment. It is thought to code for
the early antigen (EA-R). The precise start and end of this coding
region has not yet been determined. [3] resolved the conflicts
between [2] and [4] in favor of [4].
Location/Qualifiers
1..6011
/organism="Human herpesvirus 4"
/mol_type="genomic DNA"
/db_xref="taxon:10376"

repeat_region 1731..1768
/note="repeat copy A (last 38 bp)"

repeat_region 1769..1893
/note="repeat copy B"

repeat_region 1894..2018
/note="repeat copy C"

repeat_region 2019..2143
/note="repeat copy D"

repeat_region 2144..2268
/note="repeat copy E"

repeat_region 2269..2393
/note="repeat copy F"

repeat_region 2394..2518
/note="repeat copy G"

repeat_region 2519..2643
/note="repeat copy H"

repeat_region 2644..2768
/note="repeat copy I"

repeat_region 2769..2893
/note="repeat copy J"

repeat_region 2894..3018
/note="repeat copy K"

repeat_region 3019..3143
/note="repeat copy L"

repeat_region 3144..3168
/note="repeat copy M"

BASE COUNT 972 a 1897 c 1919 g 1223 t

ORIGIN 1 bp upstream of BamHI site.

Query Match 100.0%; Score 22; DB 14; Length 6011;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTGCCACCACGACGACCA 22
Db 538 ATTGCCACCACGACGACCA 559

RESULT 4
HHV507799 171823 bp DNA circular VRL 04-APR-2003
LOCUS Human herpesvirus 4 complete wild type genome.
DEFINITION AJ507799.1 GI:23893576
ACCESSION AJ507799.1 GI:23893576
VERSION complete genome.
KEYWORDS Human herpesvirus 4 (Epstein-Barr virus)
SOURCE Human herpesvirus 4
ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1
Arrand, J.R., Rymo, L., Walsh, J.E., Bjorck, E., Lindahl, T. and
Griffin, B.E.
Molecular cloning of the complete Epstein-Barr virus genome as a
set of overlapping restriction endonuclease fragments
Nucleic Acids Res. 9 (13), 2999-3014 (1981)
82014887

PUBMED 6269068
 REFERENCE
 AUTHORS
 TITLE
 POSSIBLE role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes
 JOURNAL Nucleic Acids Res. 9 (20), 5233-5262 (1981)
 MEDLINE 82059504
 PUBMED 7301588
 REFERENCE
 AUTHORS
 TITLE
 Deininger, P.L., Bankier, A., Farrell, P., Baer, R. and Barrell, B. Sequence analysis and in vitro transcription of portions of the Epstein-Barr virus genome
 JOURNAL V Cell. Biochem. 19 (3), 267-274 (1982)
 MEDLINE 83109311
 PUBMED 6296170
 REFERENCE
 AUTHORS
 TITLE
 Farrell, P.J., Deininger, P.L., Bankier, A. and Barrell, B. Homologous upstream sequences near Epstein-Barr virus promoters
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)
 MEDLINE 83169725
 PUBMED 6300857
 REFERENCE
 AUTHORS
 TITLE
 Bankier, A.T., Deininger, P.L., Farrell, P.J. and Barrell, B.G. Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8 Epstein-Barr virus
 JOURNAL Mol. Biol. Med. 1 (1), 21-45 (1983)
 MEDLINE 85035713
 PUBMED 6092825
 REFERENCE
 AUTHORS
 TITLE
 Seguin, C., Farrell, P.J. and Barrell, B.G. DNA sequence and transcription of the BamHI fragment B region of B95-8 Epstein-Barr virus
 JOURNAL Mol. Biol. Med. 1 (3), 369-392 (1983)
 MEDLINE 85060424
 PUBMED 6094553
 REFERENCE
 AUTHORS
 TITLE
 Jeang, K.T. and Hayward, S.D. Organization of the Epstein-Barr virus DNA molecule. III. Location of the P3HR-1 deletion junction and characterization of the NotI repeat units that form part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
 JOURNAL Virology 48 (1), 135-148 (1983)
 MEDLINE 83294686
 PUBMED 6310141
 REFERENCE
 AUTHORS
 TITLE
 Bankier, A.T., Deininger, P.L., Satchwell, S.C., Baer, R., Farrell, P.J. and Barrell, B.G. DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences
 JOURNAL Mol. Biol. Med. 1 (4), 425-445 (1983)
 MEDLINE 85060428
 PUBMED 6094955
 REFERENCE
 AUTHORS
 TITLE
 Farrell, P.J., Bankier, A., Seguin, C., Deininger, P. and Barrell, B.G. Latent and lytic cycle promoters of Epstein-Barr virus
 JOURNAL EMBO J. 2 (8), 1331-1338 (1983)
 MEDLINE 20331131
 PUBMED 10872327
 REFERENCE
 AUTHORS
 TITLE
 Jones, M.D., Foster, L., Sheedy, T. and Griffin, B.E. The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus
 JOURNAL EMBO J. 3 (4), 813-821 (1984)
 MEDLINE 84207939
 PUBMED 6327290
 REFERENCE
 AUTHORS
 TITLE
 Biggin, M., Farrell, P.J. and Barrell, B.G. Transcription and DNA sequence of the BamHI I fragment of B95-8 Epstein-Barr virus
 JOURNAL EMBO J. 3 (5), 1083-1090 (1984)
 MEDLINE 84236104
 PUBMED 6203743

REFERENCE
 AUTHORS
 TITLE
 Yates, J., Warren, N., Reisman, D. and Sugden, B. A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)
 MEDLINE 84222045
 PUBMED 6328526
 REFERENCE
 AUTHORS
 TITLE
 Gibson, T., Stockwell, P., Ginsburg, M. and Barrell, B. Homology between two EBV early genes and HSV ribonucleotide reductase and 38k genes
 JOURNAL Nucleic Acids Res. 12 (12), 5087-5099 (1984)
 MEDLINE 84247360
 PUBMED 6330697
 REFERENCE
 AUTHORS
 TITLE
 Baer, R.J., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G.F., Hudson, G.S., Satchwell, S.C., Seguin, C., Tufnell, P.S. and Barrell, B.G. DNA sequence and expression of the B95-8 Epstein-Barr virus genome
 JOURNAL Nature 310 (5974), 207-211 (1984)
 MEDLINE 84270667
 PUBMED 6087149
 REFERENCE
 AUTHORS
 TITLE
 Bodescot, M. and Perricaudet, M. Clustered alternative splice sites in Epstein-Barr virus RNAs
 JOURNAL Nucleic Acids Res. 15 (14), 5887 (1987)
 MEDLINE 87289053
 PUBMED 3039467
 REFERENCE
 AUTHORS
 TITLE
 Laux, G., Perricaudet, M. and Farrell, P.J. A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome
 JOURNAL EMBO J. 7 (3), 769-774 (1988)
 MEDLINE 88283646
 PUBMED 2840285
 REFERENCE
 AUTHORS
 TITLE
 Parker, B.D., Bankier, A., Satchwell, S., Barrell, B. and Farrell, P.J. Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B95-8 deletion region
 JOURNAL Virology 179 (1), 339-346 (1990)
 MEDLINE 91021036
 PUBMED 2171209
 REFERENCE
 AUTHORS
 TITLE
 Hatfull, G.F., Barrell, B.G., Quinn, J. and McGeoch, D. Unpublished
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 Blane, U.K., Amon, W. and Farrell, P.J. Induction of Epstein-Barr virus late promoters on small plasmids in the EBV late lytic cycle requires ori lyt
 JOURNAL Unpublished
 REFERENCE
 AUTHORS
 TITLE
 Farrell, P.J. Direct Submission
 JOURNAL Submitted (01-AUG-2002) Farrell P., Ludwig Institute for Cancer Research, Imperial College School of Medicine, St. Mary's Campus, Norfolk Place London W2 1PG
 MEDLINE
 PUBMED
 COMMENT
 This sequence was assembled from B95-8 EBV [14] and Raji EBV [18] with sequence corrections [16, 19]. The number of major internal repeat units has been reduced from 11.6 [14] to a more typical 7.6 and the B95-8 deletion sequences have been restored to give a sequence more representative of wild type EBV.
 Numbering
 like the modified B95-8 sequence [14, 16] accession number V01555, this sequence starts 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGATTC).
 Location/Qualifiers
 1..171823
 /organism="Human herpesvirus 4"
 /mol_type="genomic DNA"
 /strain="B95-8"

FEATURES
 source

```

source
    /db_xref="taxon:10376"
    join(1, .139223,151555, .171823)
    /organism="Human herpesvirus 4"
    /mol_type="genomic DNA"
    /strain="B95-8"
    /db_xref="taxon:10376"
    139224, .151554
    /organism="Human herpesvirus 4"
    /mol_type="genomic DNA"
    /strain="Raj1"
    /db_xref="taxon:10376"
    join(166103, .171823,1, .1682)
    /gene="LMP2"
    join(166103, .166458,58, .272,360, .458,540, .788,871, .951,
    1026, .1198,1280, .1495,1574, .1680)
    /gene="LMP2"
    /codon_start=1
    /product="terminal protein LMP2A"
    /protein_id="CAD53382.1"
    /db_xref="GI:23893577"
    /translation="MGSLFVPMGAGPPSPGDPDGYDGMNQYBSASSGNTPTP
    PNDERSNEPPPPYDPYNGNDHSDYQPTQDSYLGLOHNDGDPPEYS
    PRDSSQHTIEAGRSNVPCLPVIAPYLFMAIASCFTASVTVATGALS
    LLTAAVASSYAAQRLKLPVTVIAVTFEACITWRIDEPNSLFPALAAAG
    LQGIYVMTVLLIAYRRWRRLVCGGIMFACVLIIVDAVLQSPILGAVTV
    MTLILAFLVMTSPGIGTIGALLFLAALALALIGLITLTMPLMLMTLV
    VLLICSSCSGSKILLARFLYALALLAALAGSLIQTNPNSSTFRINL
    FCMLLIVAGLITLITMGSGNRYGVFMCGLLTMAGVMTVMSTLSA
    MTLAGLIFLIGALFVIRCCYCYCLTLESERPPTFRNTV"
    58, .272
    /gene="LMP2"
    /number=2
    join(59, .272,360, .458,540, .788,871, .951,1026, .1196,
    1280, .1495,1574, .1680)
    /gene="LMP2"
    /codon_start=1
    /product="terminal protein LMP2B"
    /protein_id="CAD53383.1"

exon
    /number=2
    join(59, .272,360, .458,540, .788,871, .951,1026, .1196,
    1280, .1495,1574, .1680)
    /gene="LMP2"
    /codon_start=1
    /product="terminal protein LMP2B"
    /protein_id="CAD53383.1"

Query Match 100.0%; Score 22; DB 14; Length 171823;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 ATTGCCACACACACACACCA 22
Db 37097 ATTGCCACACACACACCA 37118

RESULT 5
EBV 172281 bp DNA circular VRL 20-SEP-1999
DEFINITION Epstein-Barr virus (EBV) genome, strain B95-8.
ACCESSION V01555.1 GI:59974
KEYWORDS DNA polymerase; EBNA; genome; ribonucleotide reductase; tandem
repeat; terminal repeat.
SOURCE Human herpesvirus 4 (Epstein-Barr virus)
ORGANISM Human herpesvirus 4
VIRUSES; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 172281)
AUTHORS Arand,J.R., Rymo,L., Walsh,J.B., Bjorck,B., Lindahl,T. and
Griffin,B.E.
TITLE Molecular cloning of the complete Epstein-Barr virus genome as a
set of overlapping restriction endonuclease fragments
JOURNAL Nucleic Acids Res. 9 (13), 2993-3014 (1981)
MEDLINE 8201487
PUBMED 6269068
REFERENCE 2 (bases 1 to 172281)
AUTHORS Kozak,M.
TITLE Possible role of flanking nucleotides in recognition of the AUG
initiator codon by eukaryotic ribosomes
JOURNAL Nucleic Acids Res. 9 (20), 5233-5262 (1981)

MEDLINE 8205904
PUBMED 7301588
REFERENCE 3 (bases 1 to 172281)
AUTHORS Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.
TITLE Sequence analysis and in vitro transcription of portions of the
Epstein-Barr virus genome
JOURNAL U. Cell. Biochem. 19 (3), 267-274 (1982)
MEDLINE 83109311
PUBMED 6296170
REFERENCE 4 (bases 1 to 172281)
AUTHORS Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.
TITLE Homologous upstream sequences near Epstein-Barr virus promoters
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)
MEDLINE 83169725
PUBMED 6300857
REFERENCE 5 (bases 142687 to 159853)
AUTHORS Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G.
TITLE Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8
Epstein-Barr virus
JOURNAL Mol. Biol. Med. 1 (1), 21-45 (1983)
MEDLINE 85035713
PUBMED 6092825
REFERENCE 6 (bases 112620 to 125316)
AUTHORS Seguin,C., Farrell,P.J. and Barrell,B.G.
TITLE DNA sequence and transcription of the BamHI fragment B region of
B95-8 Epstein-Barr virus
JOURNAL Mol. Biol. Med. 1 (3), 369-392 (1983)
MEDLINE 85060424
PUBMED 6094953
REFERENCE 7 (bases 45644 to 52450)
AUTHORS Jeang,K.T. and Hayward,S.D.
TITLE Organization of the Epstein-Barr virus DNA molecule. III. Location
of the p3HR-1 deletion junction and characterization of the NotI
repeat units that form part of the template for an abundant
12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript
JOURNAL U. Virol. 48 (1), 135-148 (1983)
MEDLINE 83294686
PUBMED 6310141
REFERENCE 8 (bases 159853 to 172281)
AUTHORS Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J.
and Barrell,B.G.
TITLE DNA sequence analysis of the EcoRI DheI fragment of B95-8
Epstein-Barr virus containing the terminal repeat sequences
JOURNAL Mol. Biol. Med. 1 (4), 425-445 (1983)
MEDLINE 85060428
PUBMED 6094955
REFERENCE 9 (bases 1 to 172281)
AUTHORS Farrell,P.J., Bankier,A., Seguin,C., Deininger,P. and Barrell,B.G.
TITLE Latent and lytic cycle promoters of Epstein-Barr virus
JOURNAL EMBO J. 2 (8), 1331-1338 (1983)
MEDLINE 20331131
PUBMED 10872327
REFERENCE 10 (bases 45415 to 52824)
AUTHORS Jones,M.D., Foster,L., Shady,T. and Griffin,B.E.
TITLE The EB virus genome in Daudi Burkitt's lymphoma cells has a
deletion similar to that observed in a non-transforming strain
(p3HR-1) of the virus
JOURNAL EMBO J. 3 (4), 813-821 (1984)
MEDLINE 84207939
PUBMED 6327290
REFERENCE 11 (bases 87650 to 92703)
AUTHORS Biggin,M., Farrell,P.J. and Barrell,B.G.
TITLE Transcription and DNA sequence of the BamHI L fragment of B95-8
Epstein-Barr virus
JOURNAL EMBO J. 3 (5), 1083-1090 (1984)
MEDLINE 84236104
PUBMED 6203743
REFERENCE 12 (bases 7315 to 9312)
AUTHORS Yates,J., Warren,N., Reisman,D. and Sugden,B.
TITLE A cis-acting element from the Epstein-Barr viral genome that
permits stable replication of recombinant plasmids in latently
infected cells
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)

```

MEDLINE 84222045
 PUBMED 6328526
 REFERENCE 13 (bases 76089 to 79808)
 AUTHORS Gibson, T., Stockwell, P., Ginsburg, M. and Barrell, B.
 TITLE Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes
 JOURNAL Nucleic Acids Res. 12 (12), 5087-5099 (1984)
 MEDLINE 84247360
 PUBMED 6330697
 REFERENCE 14 (bases 1 to 172281)
 AUTHORS Baer, R.J., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G.F., Hudson, G.S., Satchwell, S.C., Seguin, C., Tufnell, P.S. and Barrell, B.G.
 TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome
 JOURNAL Nature 310 (5974), 207-211 (1984)
 MEDLINE 84270667
 PUBMED 6087149
 REFERENCE 15 (bases 1 to 172281)
 AUTHORS Bodessoc, M. and Perricaudet, M.
 TITLE Clustered alternative splice sites in Epstein-Barr virus RNAs
 JOURNAL Nucleic Acids Res. 15 (14), 5887 (1987)
 MEDLINE 87289053
 PUBMED 3039467
 REFERENCE 16 (bases 1 to 172281)
 AUTHORS Laux, G., Perricaudet, M. and Farrell, P.J.
 TITLE A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome
 JOURNAL EMBO J. 7 (3), 765-774 (1988)
 MEDLINE 88283646
 PUBMED 2840285
 REFERENCE 17 (bases 1 to 172281)
 AUTHORS Hatfull, G.F., Barrell, B.G., Quinn, J. and McGeoch, D.
 TITLE Unpublished
 JOURNAL 18 (bases 1 to 172281)
 REFERENCE Farrell, P.J. and Barrell, B.G.
 TITLE Direct Submission
 JOURNAL Submitted (05-JUN-1984)
 REFERENCE 19 (bases 1 to 172281)
 AUTHORS Farrell, P.J.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAR-1988)
 REFERENCE Farrell, P., Ludwig, Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG

COMMENT

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam HI fragment in which they start. eg BAPF1 is the third leftward frame starting in Bam HI fragment A. BOPF1 is the third rightward frame in Bam HI fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITES OF POLYA SIGNALS

This feature lists all occurrences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polya addition site. The rarely used homolog AATGAA is only listed when it is found in a position close to the end of a major reading frame.

SITES OF DONOR AND ACCEPT SEQUENCES

This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in

FEATURES

SOURCE

mRNA

Query Match 100.0%; Score 22; DB 14; Length 172281;
 Best Local Similarity 100.0%; Pred. No. 0.75;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCCACGACGACGACGACCA 22

Db 49385 ATGGCCACGACGACGACGACCA 49406

RESULT 6

HS4B958RAJ

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case. Restriction enzyme sites. Only the positions of the sites Bam HI (BAM) are listed. RPT This feature is used to define repetitive sequences. SITE DEL This feature defines deletions in B95-8 with respect to other strains such as Raji and also to deletions in other strains such as BJH1 and Daudi with respect to B95-8. SITE HPN Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurrences noted. ORGRL Denotes the region that encompasses an origin of replication (ori p) [13]. NUMBERING The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCACTCTT. To avoid renumbering the entire sequence, position 1 has been moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC). Location/Qualifiers

1. 172281

/organism="Human herpesvirus 4"

/mol_type="Genomic DNA"

/strain="B95-8"

/db_xref="taxon:10376"

58..272

mRNA

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

Qy

Db

RESULT 6

HS4B958RAJ

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

transcriptional initiation site
Proc. Natl. Acad. Sci. U.S.A. 88 (14), 6343-6347 (1991)
MEDLINE
91296817
PUBMED
1648738
REFERENCE
4 (bases 1 to 184113)
Jenson, H.B.
TITLE
Genbank Curator Program
JOURNAL
Unpublished (1992)
COMMENT
Original source text: Human herpesvirus 4 DNA.
The B95-8 genome (V01555) has a large deletion in the right side of
the genome which has been sequenced in Raji (M35547). These
sequences have been joined to form an extended and more complete,
although artifactual, EBV sequence.
For features, refer to feature tables of V01555 and M35547.

FEATURES
SOURCE
1. 184113
/organism="Human herpesvirus 4"
/mol_type="genomic DNA"
/db_xref="taxon:10376"
1. 152008
/note="B95-8 sequences (corresponds to 1-152,008 of
V01555)"
misc_feature
152009..152012
/note="Overlap of B95-8 and Raji sequences at B95-8
deletion point (corresponds to 152,009-152,012 in V01555,
and 1-4 in M35547)"
misc_feature
153013..163839
/note="Raji sequences (corresponds to 5-11,831 of M35547)"
163840..163843
/note="Overlap of B95-8 and Raji sequences at B95-8
deletion point (corresponds to 152,009-152,012 of V01555,
and 11,832-11,835 of M35547)"
misc_feature
163844..184113
/note="B95-8 sequences (corresponds to 152,013-172,282 of
V01555)"

BASE COUNT 36002 a 55824 c 54622 g 37665 t
ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 184113;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 ATGCGACGACGACGACGACCA 22
Db 49385 ATGCGACGACGACGACGACCA 49406

RESULT 7
S71027 266 bp DNA linear VRL 22-APR-2003
LOCUS
DEFINITION
nuclear protein EBNA-2 [Epstein-Barr virus type 1 EBV-L,
HIV-infected patient A, genomic Mutant, 266 nt].
S71027
ACCESSION
S71027.1 GI:545646
VERSION
S71027.1 GI:545646
KEYWORDS
Human herpesvirus 4 type 1 (Epstein-Barr virus type 1)
Human herpesvirus 4 type 1
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
1 (bases 1 to 266)
Buisson, M., Morand, P., Genoulaz, O., Bourgeat, M.J., Micoud, M. and
Seigneurin, J.M.
Changes in the dominant Epstein-Barr virus type during human
immunodeficiency virus infection
J. Gen. Virol. 75 (Pt 2), 431-437 (1994)

TITLE
JOURNAL
MEDLINE
PUBMED
REMARK
Genbank staff at the National Library of Medicine created this
entry [NCBI gisbseq 149519] from the original journal article.
This sequence comes from Fig. 1.
FEATURES
1. 266
location/Qualifiers
/organism="Human herpesvirus 4 type 1"

/mol_type="genomic DNA"
/db_xref="taxon:36352"
<1..266
/gene="EBNA-2"
CDS
1..266
/gene="EBNA-2"
/note="This sequence comes from Fig. 1"
/codon_start=2
/product="nuclear protein EBNA-2"
/protein_id="AAP19640.1"
/db_xref="GI:30048948"
/translation="HGGAPVPLCPHYRYKGLNPNPHHHHACISLSCHMVP
QSMHPLRHSIRNDSDSPERSPTVFNIPWLPSPQLPPAA"

BASE COUNT 74 a 107 c 31 g 54 t
ORIGIN

Query Match 90.9%; Score 20; DB 14; Length 266;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 ATGCGACGACGACGACGAC 20
Db 247 ATGCGACGACGACGACGAC 266

RESULT 8
BV031762 650 bp DNA linear STS 30-MAY-2003
LOCUS
DEFINITION
S21P6249FC9.T0 CZECHII/E1 Mus musculus STS genomic, sequence
tagged site.
BV031762
ACCESSION
BV031762.1 GI:31115657
VERSION
BV031762.1 GI:31115657
KEYWORDS
STS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 650)
Wade, C.M., Kulbokas, E.J., III, Kirby, A.W., Zody, M.C., Mullikin, J.C.,
Lander, E.S., Lindblad-Toh, K. and Daly, M.J.
The mosaic structure of variation in the laboratory mouse genome
Nature 420 (6915), 574-578 (2002)
JOURNAL
MEDLINE
PUBMED
22354684
12466852

COMMENT
Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersli@genome.wi.mit.edu
Primer A: None
Primer B: None
STS size: 650
Protocol:
WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 125S1/SVIMU, C3H/HeJ, and B6H/cByJ. The WGS
reads were placed uniquely on the MGSVC3 C57BL/6J assembly and SNP
detection was carried out by SSAHA-SNP. 225,000 reads were
annotated
as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
and the strain from which the particular read came. The validation
rate for these SNPs was estimated at approximately 98%.
FEATURES
SOURCE
1. 650
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CZECHII/E1"
/db_xref="taxon:10090"
/map="1 21-604 145737352-145737935"
/clone_lib="CZECHII/E1"
STS
1..650
/organism="Human herpesvirus 4 type 1"

BASE COUNT 182 a 141 c 117 g 210 t

ORIGIN

Query Match 90.3%; Score 20; DB 11; Length 650;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3 TGCACACACGACGACCA 22
 568 TGCACACACGACGACCA 587

RESULT 9
 AC009695/c 144328 bp DNA linear PRI 23-JUL-2002
 LOCUS Homo sapiens chromosome 8, clone RP11-130C12, complete sequence.
 DEFINITION AC009695
 VERSION AC009695.7 GI:21930259
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 144328)
 Homo sapiens chromosome 8, clone RP11-130C12
 Unpublished
 2 (bases 1 to 144328)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Baker, J., Baldwin, J., Barna, N., Beckwith, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., DeRubeis, K., DePuy, E., Devon, K., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
 Funk, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
 Hages, B., Heath, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karatas, A., Lehoczy, J., Lien, C., Locke, K., MacDonald, P.,
 Marcus, N., McEwan, P., McGurk, A., McKernan, K., McDonald, J.,
 Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Nigam, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollard, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Tesfaye, S., Torrella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
 Direct Submission
 Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 144328)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Bouhagalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Chazargal, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., DeRubeis, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Grant, G.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kelle, C., LaRoque, K., Lamares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., MacDonald, P., Major, J., Margis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollard, V.,
 Raymond, C., Reta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Struss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 144328)

AUTHORS
 Birren, B., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhagalter, B.,
 Camarata, J., Chang, J., Chazargal, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., DeRubeis, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hages, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karatas, A., Kelle, C., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mlenga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Reta, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Struss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (23-JUL-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 23, 2002 this sequence version replaced gi:19881988.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WtBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 12317
 Center clone name: 130_C12.

 Only the last 144.3 kilobases of this clone are being submitted.
 The remaining overlaps accession number AC022716 (WIGC Project
 15398).

 FEATURES
 source
 1. 144328
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-130C12"
 /clone_lib="RP11-130C12 Human Male BAC"
 202. 628
 /rpt_family="MLT1C"
 926. 1461
 /rpt_family="HERV16"
 complement(1498. 1845)
 /rpt_family="MLT1A1"
 1893. 2312
 /rpt_family="HERV16"
 2348. 2811
 /rpt_family="HERV16"
 2853. 3399
 /rpt_family="HERV16"
 3404. 3506
 /rpt_family="LTR133"
 3654. 3721
 /rpt_family="CATGn"
 complement(4271. 4466)
 /rpt_family="L2"
 4967. 5065
 /rpt_family="MIR"
 complement(5179. 5473)
 /rpt_family="ALUSx"
 complement(5522. 5721)
 /rpt_family="MLT1H"
 5797. 5891
 /rpt_family="MLT1G1"

```

repeat_region      complement(5922, .6218)
                    /rpt_family="MLT1H"
repeat_region      complement(6219, .6550)
                    /rpt_family="MLT1H-int"
repeat_region      complement(7546, .7633)
                    /rpt_family="MLT1H-int"
repeat_region      complement(7640, .8064)
                    /rpt_family="MLT1H"
repeat_region      8429, .8614
                    /rpt_family="MLT1H"
repeat_region      complement(8781, .8912)
                    /rpt_family="L2"
repeat_region      complement(8966, .9117)
                    /rpt_family="MARNA"
repeat_region      complement(10658, .10704)
                    /rpt_family="L2"
repeat_region      complement(11250, .11297)
                    /rpt_family="AluSg"
repeat_region      complement(11298, .11588)
                    /rpt_family="AluCb"
repeat_region      complement(11589, .11652)
                    /rpt_family="AluSg"
repeat_region      complement(11677, .11972)
                    /rpt_family="AluSg"
repeat_region      11984, .12079
                    /rpt_family="CT-rich"
repeat_region      13713, .13821
                    /rpt_family="L2"
repeat_region      14617, .14765
                    /rpt_family="MIR"
repeat_region      complement(15582, .15948)
                    /rpt_family="MLT1A1"
repeat_region      16797, .16821
                    /rpt_family="AT-rich"
repeat_region      17352, .17512
                    /rpt_family="MER5A"
repeat_region      complement(17529, .18010)
                    /rpt_family="Char164a"
repeat_region      complement(18558, .18686)
                    /rpt_family="MER103"
repeat_region      complement(18687, .18739)
                    /rpt_family="MER5B"
repeat_region      18740, .19050
                    /rpt_family="AluSg"
repeat_region      complement(19051, .19116)
                    /rpt_family="MER5B"
repeat_region      complement(19190, .19551)
                    /rpt_family="MLT1C"
repeat_region      complement(19552, .19713)
                    /rpt_family="AluSx"
repeat_region      19714, .19744
                    /rpt_family="TG.n"
repeat_region      complement(19745, .19866)
                    /rpt_family="AluSx"
repeat_region      complement(19867, .19968)
                    /rpt_family="MLT1C"
repeat_region      complement(20532, .20582)
                    /rpt_family="MIR"
repeat_region      20605, .20677
                    /rpt_family="MIR"
repeat_region      21274, .21569
                    /rpt_family="AluSc"
repeat_region      22137, .22284
                    /rpt_family="MIR"
repeat_region      23081, .23224
                    /rpt_family="TA.n"
repeat_region      23898, .23928
                    /rpt_family="(CAAAA)n"

```

Query Match 90.9%; Score 20; DB 9; Length 144328;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      3  TGGCACCACGACGACCA  22
Db      29100  TGGCACCACGACGACCA  29081

RESULT 10
LOCUS      AF247362
DEFINITION Petromyzon marinus Spi transcription factor mRNA, complete cds.
ACCESSION  AF247362
VERSION    AF247362.1  GI:8745403
KEYWORDS
SOURCE
ORGANISM   Petromyzon marinus (sea lamprey)
            Petromyzon marinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Heteroartia;
            Petromyzontiformes; Petromyzontidae; Petromyzon.
REFERENCE  1 (bases 1 to 1272)
            Shintani,S., Terzic,J., Sato,A., Saraga-Babic,M., O'Nigin,C.,
            Tichy,H. and Klein,J.
            Do lampreys have lymphocytes? The Spi evidence
            Proc. Natl. Acad. Sci. U.S.A. 97 (13), 7417-7422 (2000)
            MEDLINE 20319041
            PUBMED  10840049
REFERENCE  2 (bases 1 to 1272)
            Shintani,S.
            Direct Submission
            Submitted (22-MAR-2000) Abteiling Immunogenetik,
            Max-Planck-Institut fur Biologie, Corrensstr. 42, Tubingen
            D-72076, Germany
FEATURES
     source          1..1272
                     /organism="Petromyzon marinus"
                     /mol_type="mRNA"
                     /db_xref="taxon:7757"
                     224..1105
                     /codon_start=1
                     /product="Spi transcription factor"
                     /protein_id="AF247362.1"
                     /db_xref="GI:8745404"
                     /translation="MLAMHTQPMYATLRLDALASGYDEASTLYSSDTNLSKYDPL
                     DIIDYITHELQKCGNSDKNPISGSDVLTSGEGSSSYCSLLDEPHAPVSPPTLS
                     SADADYSDVPTLQISDTSEDEEESEEDBECCRCRPVLEPFGVAGGPGGQ
                     OOOOGRRHHQHHQOOOOLPQGVYKXKKGXKRLVQFLWEIIDDQRMCH
                     CIMWDEROGIFGSSSCHKEELAKKMGKKNRRAMTYQKARLRALRYEATGIRKIK
                     KKUTYQFARKMGAFRR"
BASE COUNT      276 a      384 c      409 g      203 t
ORIGIN
Query Match 86.4%; Score 19; DB 5; Length 1272;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4  GCCACCCACGACGACCA  22
Db      729  GCCACCCACGACGACCA  747

RESULT 11
LOCUS      AX078616
DEFINITION Sequence 130 from Patent WO0107624.
ACCESSION  AX078616
VERSION    AX078616.1  GI:13158258
KEYWORDS
SOURCE
ORGANISM   Pseudomonas putida
            Pseudomonas putida
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1
            Fraser,C.M., Venter,C., Tuemler,B., Hohelsel,J., Duesterhoeft,A.,
            Hilbert,H., Timms,K.N., Moore,E., Straetz,M., Heim,S.,
            Nelson,K.E., Hickey,E. and Peterson,J.
            Dna sequences which are suited for specifically detecting

```

JOURNAL Patent: WO 0107624-A 130 01-FEB-2001;
THE INSTITUTE FOR GENOMIC RESEARCH (US) ; QIAGEN GmbH (DE) ;
Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE) ;
Deutsches Krebsforschungszentrum (DKFZ) (DE) ; Medizinische
Hochschule Hannover (DE)

FEATURES
source
1. .2001
/organism="Pseudomonas putida"
/mol_type="genomic DNA"
/db_xref="taxon:303"

BASE COUNT 333 a 719 c 619 g 325 t 5 others

Query Match 86.4%; Score 19; DB 6; Length 2001;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCCAGCAGCAGCACCA 22
Db 1001 GCCACCCAGCAGCAGCACCA 1019

RESULT 12
LOCUS AX078615 2028 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 129 from Patent WO0107624.
ACCESSION AX078615
VERSION AX078615.1 GI:13158257

SOURCE
Pseudomonas putida
Pseudomonas putida
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE
AUTHORS Fraser,C.M., Venter,C., Tuemmler,B., Hohnesel,J., Duesterhoeft,A.,
Hilbert,H., Jimmis,K.N., Moore,E., Straetz,M., Heim,S.,
Nelson,K.E., Hickey,E. and Peterson,J.
Dna sequences which are suited for specifically detecting
Pseudomonas putida Kc2440
Patent: WO 0107624-A 129 01-FEB-2001;
THE INSTITUTE FOR GENOMIC RESEARCH (US) ; QIAGEN GmbH (DE) ;
Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE) ;
Deutsches Krebsforschungszentrum (DKFZ) (DE) ; Medizinische
Hochschule Hannover (DE)

JOURNAL
Pseudomonas putida Kc2440
Patent: WO 0107624-A 129 01-FEB-2001;
THE INSTITUTE FOR GENOMIC RESEARCH (US) ; QIAGEN GmbH (DE) ;
Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE) ;
Deutsches Krebsforschungszentrum (DKFZ) (DE) ; Medizinische
Hochschule Hannover (DE)

FEATURES
source
1. .2028
/organism="Pseudomonas putida"
/mol_type="genomic DNA"
/db_xref="taxon:303"

BASE COUNT 331 a 628 c 710 g 356 t 3 others

Query Match 86.4%; Score 19; DB 6; Length 2028;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCCAGCAGCAGCACCA 22
Db 869 GCCACCCAGCAGCAGCACCA 851

RESULT 13
LOCUS AF109781 2847 bp mRNA linear VRT 20-APR-1999
DEFINITION Danio rerio basic domain leucine zipper transcription factor (maf)

ACCESSION AF109781
KEYWORDS AF109781.1 GI:4588533

SOURCE
Danio rerio (zebrafish)
Danio rerio

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 2847)
Schwarzstein,M., Kirm,A., Haffter,P. and Cordes,S.P.
Expression of Zkrm12, a homologue of the Krm11/val segmentation
gene, during embryonic patterning of the zebrafish (Danio rerio)
Mech. Dev. 80 (2), 223-226 (1999)

JOURNAL MEDLINE 99173794
PubMed 10072793

REFERENCE
AUTHORS Schwarzstein,M. and Cordes,S.P.
TITLE Direct Submission
JOURNAL Submitted (26-NOV-1998) Samuel Lunenfeld Research Institute, Mt.
Sinai Hospital, 600 University Ave, Toronto, ON M5G 5X1, Canada

FEATURES
source
1. .2847
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"

gene
1. .2847
/gene="maf"
302. 1276
/gene="maf"
/codon_start=1
/product="basic domain leucine zipper transcription
factor"
/protein_id="AA026141.1"
/db_xref="GI:4588534"

BASE COUNT 795 a 635 c 588 g 824 t 5 others

Query Match 86.4%; Score 19; DB 5; Length 2847;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCCAGCAGCAGCACCA 22
Db 846 GCCACCCAGCAGCAGCACCA 864

RESULT 14
LOCUS AF333961 2981 bp DNA linear BCT 01-MAY-2001
DEFINITION Pseudomonas putida Cadr (cadr) and Cadr (cadr) genes, complete cds;
and lysr family response regulator (lysR) gene, partial cds.

ACCESSION AF333961
VERSION AF333961.1 GI:13898624
KEYWORDS
Pseudomonas putida
Pseudomonas putida
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE
AUTHORS Lee,S.W., Gluckman,E. and Cooksey,D.A.
TITLE Chromosomal locus for cadmium resistance in Pseudomonas putida
consisting of a cadmium-transporting ATPase and a Werr family
response regulator
Appl. Environ. Microbiol. 67 (4), 1437-1444 (2001)

JOURNAL MEDLINE 2178487
PubMed 1182558

REFERENCE
AUTHORS Lee,S.W., Gluckman,E. and Cooksey,D.A.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2001) Plant Pathology, University of California,
Riverside, CA 92521, USA

FEATURES
source
1. .2981
Location/Qualifiers


```

/organism="Pseudomonas putida"
/mol_type="genomic DNA"
/db_xref="taxon:303"
/complement(12..455)
/gene="cadA"
/complement(12..455)
/gene="cadA"
/feature="Wear family response regulator; regulation of cadmium resistance"
/codon_start=1
/transl_table=11
/product="CadR"
/protein_id="AAK48831.1"
/db_xref="GI:13898625"
/translation="MKIGELAKATDCAVETIRYERQQLPEPARSDGNVRLYTQAHV
ERLFIINCRITLMDTDEIRSLRLRDSPPDSCGYNALIDEHIEHQARIDQLVALQ
BOVELRRRCNAGACACALQLETNAGVSPTEHSHVGRSHGH"
540..2753
/gene="cadA"
540..2753
/gene="cadA"
/feature="cadmium transporting ATPase; cadmium resistance"
/codon_start=1
/transl_table=11
/product="CadA"
/protein_id="AAK48831.1"
/db_xref="GI:13898626"
/translation="MNQPVSEHKHDMHNSCCGSAAPALVOLSEKASHAOFRRER
IEAMDCEPQTLIDKGLAGIEQLEFNINLVGLRITLGTALIEQALISLQKA
EPGAEEDSTASAPQAKTRMPLALSGAIAEIVHFPALAPENVALALIAIG
CGATYKKGWIALKNRNLINAMLSIAVGAIVIGQPEAAVMVLTFTVAELIARSL
DRRNAIGIMOLTPMATYROADGOMREVEVEVAIGALNVRPGERIGLGEATSG
OSVDAQPRIGESLPKVTYGDLPAGTINQAGALEFRVTAAGOSTLIRIKAVEEA
QKRAPIQRFVDRFSRIYFVFAIALAVAIPELFWAGAFDMVRAVLIVACPC
ALVISTPVITVSGLAARAGILIKGVYLEGRQDFTALDKTGTITGKRVQDPAK
ILBPLFEGRAQALASIGERSDHPRAIAQFEGQALSEVDDEPALAGVADTVKE
AGEVYILGNHRLVEELGCSPELEADLIERGQKTVLLDRSGPLALFAVADTVKE
SSQVLAIEHEILGKITVMLTGNPHTAQAIAATVGTIDRAEGNLLPDKKISITLYAO
GHRVAGVGDGINDAIPALAPAEIGFAMAGTQRAFTADVALMDDDKIRIPAFVLSR
QSAAILMONIVLALGKALFLAITTPAGATMMWAFADMGVSLVFNGLRLRK"
2763..2981
/gene="lyrA"
/feature="lyrA"
/feature="transcriptional regulator"
/codon_start=1
/transl_table=11
/product="LyrA family response regulator"
/protein_id="AAK48832.1"
/db_xref="GI:13898627"
/translation="MUSELKAFVWVARLGSITLAKKUGLSQPTVTTQIRLBSQVA
VELFYRGRRLVISEGVRLPVRKALQ"
947..982
515 a 982 g 537 t
BASE COUNT
ORIGIN
Query Match 86.4%; Score 19; DB 1; Length 2981;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

ORGANISM
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 3007)
AUTHORS
Kim,H.R., Yew,N.S., Anseorge,W., Voss,H., Schwager,C., Vennstrom,B.,
Zenke,M., and Engel,J.D.
Two different mRNAs are transcribed from a single genomic locus
encoding the chicken erythrocyte anion transport proteins (band 3)
JOURNAL
Mol Cell Biol 8 (10), 4416-4424 (1988)
MEDLINE
89039870
PUBMED
3185555
COMMENT
On Aug 20, 1994 this sequence version replaced gi:340910.
source text: Gallus gallus cdna to mRNA.
FEATURES
source
1..3007
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
77..2845
/codon_start=1
/product="erythrocyte anion transport protein"
/protein_id="AA48753.1"
/db_xref="GI:531173"
/translation="MEGPGQTDALRRSLDPEGYEDTKSGRISLGTMSNPLVSDVL
BAAGROPTAHRTDYEGLVELDELVLSRQDPCWMEGRLMLESMEPGAGWSHL
LITVHSLIELHRAFAKGVILDVANSLAAVAHVLIDQLIYEQLEKQHRDVLRAL
LIRHGPSAESVWTLPAALQDSCDEGADDERLALDORVEREILHGAQSPSRQ
LGPOHLOQLEDTATVILVCAAFLOPLIALRLAPGADAVLAPLPRVLTVA
SCQHLHLIPQRANRRRYHPPTVSPGPTPKYTGKQCAPODDDLRTRRF
GCVLRDTRRRYKRLSDIRDLNQCIAAVITFYFALSPAITFGALLGETKGMWV
SELLSTVQCLLSLSAQLLVVGSGLVFEAEFFPCEDHGLETVGRWVIGF
WDLIVLVVACEGTIVRYSRYTOEIFFSLISLIYEFALVTIPEAHPOQSY
DIDVSTSPSVKPMNTALSLVMAGTEFLALFLQFQNSVFLPGKVRLLIGDFVPS
IFWALADFQIDKOTYQKLVKPERKLVNGARWMPFIHPGASATPPIIMWPSVPA
LIVPILFLEQITTLIVSKPERKLVNGSGPHLIDLLIVMGALALFGMPWLSATTV
PRTIHANALTVGSGANPGERAHVYKRBQSLUYAVLTGSLIMEPLIKTIPAV
LFGFLVNGVTSGLQIFDELILMLPPKHPEPYTVKTRITSSPLUTOLVA
LIMGVKSPASLRCPVLVITPRLRLPRITSEIELKCLDIDDAVATVEEAGQV
YNEVQMPs"
BASE COUNT
ORIGIN
Query Match 86.4%; Score 19; DB 5; Length 3007;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


PCH33-1.
 Drate entry and computer-readable sequence for [1] kindly provided
 by J.V.Cox, 02-MAR-1988.
 Location/Owner

FEATURES	Location/Qualifiers
SOURCE	1. .3407

BASE COUNT	644 a	1081 c	998 g	684 t
ORIGIN	693 bp upstream of BamHI site.			

Query Match	86.4%;	Score 19;	DB 5;	Length 3407;
Best Local Similarity	100.0%;	Pred. No. 37;		
Matches 19;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 4 GCCACCCAGCAGCACC 22
Db 2027 GCCACCCAGCAGCACC 2009

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
RESULT 17	AC018248	AC018248	Drosophila melanogaster.	9355 bp	DNA	linear	HTG 09-DEC-1999
			*** SEQUENCING IN PROGRESS ***.				
			AC018248.1	GI:6552943	HTG; HTGS_PHASE2..	Drosophila melanogaster (fruit fly)	Drosophila melanogaster

REFERENCE 1 (bases 1 to 9359)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM-10214038 by the submitter
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source location/Qualifiers
1..9359

Query Match	86.4%	Score 19;	DB 2;	Length 9359;
Best Local Similarity	100.0%	Pred. No. 32;		
Matches 19; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY 4 GCCACCACCGCAGCACCA 22
Db 8036 GCCACCACCGCAGCACCA 8054

RESULT	18
LOCUS	AC142283
DEFINITION	Homo sapiens chromosome 2 clone XPOS-934649, complete sequence.
ACCSSION	AC142283
VERSION	AC142283.3 GI:29501957
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE	5 (bases 1 to 26546)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (04-APR-2003) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Apr 3, 2003 this sequence version replaced gi:293363315.

FEATURE	source	Location/Qualifiers
1..26546		
		/organism="Homo sapiens"
		/mol_type="genomic DNA"
		/db_xref="taxon:9606"
		/chromosome="2"
		/clone="XF05-83404G9"
BASE COUNT	6789 a	5979 c 6469 g 7309 t
ORIGIN		

Query Match	86.4%	Score 19	DB 9	Length 26546
Best Local Similarity	100.0%	Pred. No. 27		
Matches 19	Conservative 0	Mismatches 0	Indels 0	Gaps 0

4 GCCACCACCAGCAGCACCA 22

Mon Aug 18 10:30:10 2003

us-10-074-620-6.011.rge

Page 13

DB 8135 GCCACCACGACGACCA 8153

RESULT 19
AF452638

LOCUS AF452638 75022 bp DNA linear PRI 31-MAR-2002
DEFINITION Homo sapiens distal-less homeo box 7 (DLX7) and distal-less homeo
box 3 (DLX3) genes, complete cds.

ACCESSION AF452638
VERSION AF452638.1 GI:19849182

KEYWORDS
SOURCE
ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE

Sumiyama, K., Irvine, S.O., Stock, D.W., Weiss, K.M., Kawasaki, K.,
Shimizu, N., Shashikant, C.S., Miller, W. and Ruddle, F.H.
Genomic structure and functional control of the DLX3-7 bigene
cluster
Proc. Natl. Acad. Sci. U.S.A. 99 (2), 760-785 (2002)

JOURNAL
MEDLINE
PUBMED

2 (bases 1 to 75022)
11792834
Sumiyama, K., Irvine, S.O., Stock, D.W., Weiss, K.M., Kawasaki, K.,
Shimizu, N., Shashikant, C.S., Miller, W. and Ruddle, F.H.

TITLE
JOURNAL
SUBMITTED (29-NOV-2001) MCDB, Yale University, 266 Whitney Avenue,
New Haven, CT 06520, USA

FEATURES
source
location/Qualifiers

1..75022
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="PI-1490"
/cell_type="fibroblast"
/tissue_type="foreskin"
246..344
/note="L3b"
repeat_region
/rpt_family="LINE/CR1"
complement(1889..1986)
/note="MIR"
repeat_region
/rpt_family="SINE/MIR"
complement(4242..4442)
/note="MIR"
repeat_region
/rpt_family="SINE/MIR"
complement(5027..5164)
/note="L1ND3"
repeat_region
/rpt_family="LINE/L1"
5165..5460
/note="AluJo"
repeat_region
/rpt_family="SINE/Alu"
complement(5461..5582)
/note="L1ND3"
repeat_region
/rpt_family="LINE/L1"
7829..7983
/note="L2"
repeat_region
/rpt_family="LINE/L2"
7987..8241
/note="Zaphod"
repeat_region
/rpt_family="DNA/Achobo"
8304..8368
/note="L1McA"
repeat_region
/rpt_family="LINE/L1"
8373..8629
/note="Zaphod"
repeat_region
/rpt_family="DNA/Achobo"
complement(8630..8923)
/note="AluSg"
repeat_region
/rpt_family="SINE/Alu"
8924..9103
/note="Zaphod"

repeat_region
/rpt_family="DNA/Achobo"
9104..9405
/note="AluJo"
repeat_region
/rpt_family="SINE/Alu"
9406..9515
/note="Zaphod"
repeat_region
/rpt_family="DNA/Achobo"
9561..9859
/note="AluSg"
repeat_region
/rpt_family="SINE/Alu"
9869..10162
/note="AluSx"
repeat_region
/rpt_family="SINE/Alu"
10215..10280
/note="MER115"
repeat_region
/rpt_family="DNA/Achobo"
complement(12161..12406)
/note="MIR3"
repeat_region
/rpt_family="SINE/MIR"
complement(12883..12956)
/note="MIR"
repeat_region
/rpt_family="SINE/MIR"
13319..13438
/note="FLAM A"
repeat_region
/rpt_family="SINE/Alu"
complement(13487..13611)
/note="L2"
repeat_region
/rpt_family="LINE/L2"
complement(13626..13822)
/note="L2"
repeat_region
/rpt_family="LINE/L2"
14348..14504
/note="L3"
repeat_region
/rpt_family="LINE/CR1"
complement(14561..14728)
/note="L2"
repeat_region
/rpt_family="LINE/L2"
15132..15163
/note="(TG)n: simple repeat"
complement(15270..15578)
/note="AluJo"
repeat_region
/rpt_family="SINE/Alu"
complement(15643..15808)
/note="MIR"
repeat_region
/rpt_family="SINE/MIR"
16558..16860
/note="AluSx"
repeat_region
/rpt_family="SINE/Alu"
16861..17159
/note="AluY"
repeat_region
/rpt_family="SINE/Alu"
complement(17260..17394)
/note="L3"
repeat_region
/rpt_family="LINE/CR1"
19388..19616
/note="L2"
repeat_region
/rpt_family="LINE/L2"
complement(19617..19895)
/note="AluSg"
repeat_region
/rpt_family="SINE/Alu"
19896..20254
/note="L2"
repeat_region
/rpt_family="LINE/L2"
21045..21253
/note="MIR"
repeat_region
/rpt_family="SINE/MIR"
complement(21867..22053)
/note="MIR"
repeat_region
/rpt_family="SINE/MIR"
22410..22717
/note="AluY"
repeat_region
/rpt_family="SINE/Alu"
complement(23157..23242)

```

/note="MIR"
/rpt_family="SINE/MIR"
repeat_region
25176..25237
/note="(CA)n; simple repeat"
complement(25620..25680)
/note="L2"
/rpt_family="LINE/L2"
complement(26172..26484)
/note="AluSg"
/rpt_family="SINE/Alu"
26492..26546
/note="AT rich; low complexity"
complement(26547..26679)
/note="FLAM C"
/rpt_family="SINE/Alu"
complement(27219..27267)
/note="MIR3"
/rpt_family="SINE/MIR"
complement(27268..27566)
/note="AluSp"
/rpt_family="SINE/Alu"
complement(27567..27579)
/note="MIR3"
/rpt_family="SINE/MIR"
complement(27646..27842)
/note="MIR"
/rpt_family="SINE/MIR"
complement(27839..27912)
/note="L3"
/rpt_family="LINE/CR1"
complement(28082..28139)
/note="L3"
/rpt_family="LINE/CR1"
28404..28581
/note="MIR"
/rpt_family="SINE/MIR"
complement(28955..29097)
/note="MIR"
/rpt_family="SINE/MIR"
29103..29239
/note="AluUd"
/rpt_family="SINE/Alu"
29244..29554
/note="AluSg"
/rpt_family="SINE/Alu"
29609..29961
/note="AluSc"
/rpt_family="SINE/Alu"
30142..30276
/note="MIR"
/rpt_family="SINE/MIR"
complement(30277..30582)
/note="AluSg"
/rpt_family="SINE/Alu"
30583..30713
/note="MIR"
/rpt_family="SINE/MIR"
31885..31989
/note="(TGT)n; simple repeat"
complement(32031..32118)

```

Query Match 86.4%; Score 19; DB 9; Length 75022;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TGCACACACAGACACAC 21
 DB 6458 TGCACACACAGACACAC 6476

RESULT 20
 LMFCHR32_02/c
 WPCOMMENT

Sequence split into 28 fragments LOCUS LMFCHR32 Accession AL499622

Fragment Name	Begin	End
LMFCHR32_00	1	110000
LMFCHR32_01	100001	210000
LMFCHR32_02	200001	310000
LMFCHR32_03	300001	410000
LMFCHR32_04	400001	510000
LMFCHR32_05	500001	610000
LMFCHR32_06	600001	710000
LMFCHR32_07	700001	810000
LMFCHR32_08	800001	910000
LMFCHR32_09	900001	1010000
LMFCHR32_10	1000001	1110000
LMFCHR32_11	1100001	1210000
LMFCHR32_12	1200001	1310000
LMFCHR32_13	1300001	1410000
LMFCHR32_14	1400001	1510000
LMFCHR32_15	1500001	1610000
LMFCHR32_16	1600001	1710000
LMFCHR32_17	1700001	1810000
LMFCHR32_18	1800001	1910000
LMFCHR32_19	1900001	2010000
LMFCHR32_20	2000001	2110000
LMFCHR32_21	2100001	2210000
LMFCHR32_22	2200001	2310000
LMFCHR32_23	2300001	2410000
LMFCHR32_24	2400001	2510000
LMFCHR32_25	2500001	2610000
LMFCHR32_26	2600001	2710000
LMFCHR32_27	2700001	2727709

Continuation (3 of 28) of LMFCHR32 from base 200001 (AL499622 Leishmania major chromosome 1)

Query Match 86.4%; Score 19; DB 2; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACACACAGACACACCA 22
 DB 41219 GCCACACACAGACACCA 41201

RESULT 21
 LOCUS AP003279 134982 bp DNA linear PLN 29-MAR-2003
 DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
 PAC clone: P0529E05.
 ACCESSION AP003279 BA000010
 VERSION AP003279.2 GI:15718426
 KEYWORDS
 ORGANISM
 SOURCE
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Euphorbiaceae; Oryzae; Oryza.

REFERENCE
 AUTHORS
 1 Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
 Katayose, Y., Wu, J., Mitamura, Y., Cheng, Z., Nagamura, Y.,
 Antonio, B. A., Kanamori, H., Hosokawa, S., Masuoka, M., Aikawa, K.,
 Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
 Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y.,
 Idonuma, A., Iijima, M., Ikeda, M., Ikono, M., Itoh, S., Itoh, T.,
 Itoch, Y., Itoh, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S.,
 Kikuta, A., Kobayashi, N., Kono, I., Machida, K., Maehara, T.,
 Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,
 Nakama, Y., Nakamichi, Y., Nakamura, M., Nakai, N., Negishi, M.,
 Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, N., Shimokawa, T.,
 Shomura, A., Song, U., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,
 Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K.,
 Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,
 Yano, M., Jiang, J. and Gojobori, T.
 The genome sequence and structure of rice chromosome 1
 Nature 420 (6913), 312-316 (2002)
 22337376

PIRMBED 12447438
 2 (bases 1 to 134982)
 Sasaki,T., Matsumoto,T. and Yamamoto,K.
 AUTHORS
 Direct Submission
 JOURNAL
 Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program, Kamondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 On Sep 20, 2003 this sequence version replaced gi:13027309.
 Genes were predicted from the integrated results of the following:
 GENSSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
 (October 1998 version). The genomic sequence was searched against
 NCBI Nonredundant Protein database, nr
 (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
 RGP. Protein homologues of the coding regions were searched against
 NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent
 the identified cDNA sequences using BLASTN 2.0 with the
 corresponding DBJ accession no. and RGP clone ID.

COMMENT

A gene with identity or significant homology to a protein is
 classified based on the protein name to indicate the homology level
 such as same name, 'putative-' and '-like protein'. A gene without
 significant homology to any protein but with EST homology (covering
 almost the entire length of partial sequence) is classified as an
 'unknown' protein. A gene predicted with a gene prediction program
 is classified as a 'hypothetical' protein.
 The orientation of the sequence is from 5' to 3' of the PAC clone.
 This sequence of P0529E05 clone has an overlap with P0529H11 (DBJ:
 AP004072) clone at the position 1 to 19,018 of 5' end. The sequence
 of this clone starts at the position 146,952 of P0529H11. Detailed
 information on overlap and assembly quality together with
 annotation of this entry is available at
 http://rgp.dna.affrc.go.jp/genomeseg.html.

FEATURES

SOURCE

1. 134982
 /organism="Oryza sativa (Japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="1"
 /clone="P0529E05"
 /complement(788..1069)
 /gene="P0529E05.1"
 /complement(788..1069)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB84386.1"
 /db_xref="GI:18461193"
 /translation="MASVGGGGRVAAAPRQRAAVSATVAGAGGAGGAMARQR
 QGDNQVDEMPTRDVVVELTVYERKFLPVGNHPLVPGFPVVAIRP"
 /complement(join(3596..3889,3976..4274,4750..4839,
 5288..5456,5562..5684,6335..6361))
 /gene="P0529E05.2"
 /complement(join(3596..3889,3976..4274,4750..4839,
 5288..5456,5562..5684,6335..6361))
 /note="contains EST R03279 (R3279)"
 /hypothetical protein"
 /codon_start=1
 /protein_id="BAB84387.1"
 /db_xref="GI:18461190"
 /translation="KAMIFLLVDFICATNSRSEQRIGAGNTGYKKAADGSI
 VKATKMTDRHLSSEPRSEIQTLSKTEHLNLFGLIHEDERLILVEYNSSLR
 EHDGJLGEPLSSQRLNLAIDIVHAVSYLHGTHPIIHRDIKSNILITLQVLRKY
 EHDGJLGEPLSSQRLNLAIDIVHAVSYLHGTHPIIHRDIKSNILITLQVLRKY
 RPIERGRGHQRLITQWALRKCRDGAIVVADAMRRTSAVVAAMEKMLAAEETA
 PDAARPMARCEVLSIRDFDQCOQATAAAGARGRDGSYPSITSLKE"
 /complement(join(7920..8104,8190..8223,8332..8469,
 8893..9039,9136..9180,9283..9646,9758..9984,10424..10843))
 /gene="P0529E05.3"
 /complement(join(7920..8104,8190..8223,8332..8469,
 8893..9039,9136..9180,9283..9646,9758..9984,10424..10843))
 /note="contains EST AU090537 (C1241)"
 /unknown protein"
 /codon_start=1
 /protein_id="BAB84391.1"
 /db_xref="GI:18461194"
 /translation="MAQRDEQILYDTAEGSSGYWNEEGCNPNODLNEGVARD
 AEGNQGEMERDVNGEESGSPQVQGRAGOGAAGKXGRIITEVEDGSR
 SAPDAKNTYRSHGVVDDVPSVNYWTRTRGHSFVPSSEKEMTLITLLETF
 TLPAETEKYRGTWKLMKAOEFOSEKELLYKTLKOTPTFDPFPLRLDMDFFVY
 KTGBOGAMMERKNENAKKRYHHHLSGGVSYVMPMESEATPDRGEPAANP
 ERSKPYVAGNCTLNPAADGIVEDQVPRBAARLTNAVBSAGSTPDRGEPAANP
 IQPHEHGOGRGKAVIPWKIGFKEDITHTYSQKSKDTEKADLEFRVSSYELNQ
 EVARVDERMARSNPPOTIPPAVSPGNSGASTQVGSQMDMWTQDEST
 CPVDDITQWTCLEHLPKSLIKVAGMAIPTDPGQTHCRPIPVGSKVEVELVVG

/gene="P0529E05.3"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB84388.1"
 /db_xref="GI:18461191"
 /translation="WAPPGQLPLARSLLPLSAPPVSGRRRLPTLVIGRALPPPT
 LPHGLSPHPLPFPAPPRRLSRPPPATSLPGASPGGAQEAQVAAEFVTSRVVAA
 MGLALACNADRVVMSVAIVLSQAVGWSVQVSSFLMGVLPPIIGALVAVY
 YGKVMVAVGVALMSLATPLSPMAASLWLEPSTRVLLGMAEGVALPSNMNVLMF
 PTERSSAVGIMAGFQGLITGLILSIRRAIIPQPVIFGLVGLVWLTSAI
 SCTPEMNOISAEHEDYTRKQKYKTSQSGERLAKVPPTSLSKMTPLISANM
 HSNQFVLSKMPVYFKTITHNLNEMASLPMVMAVGVYVGVSDRLIQGTS
 ILYTRKIQTLGFGVGVALLGNLAKSPVASAWLTAVGLKSGHSGFLVNLQEIA
 PQYAVLVHGMNSTAGTFALIGTVAGGFPVDRMGSFRGLITLSLYPSTLFWDIFA
 TGERVDPDGTG"
 /complement(join(14141..14355,14614..14942,15117..15432,
 15672..15826,15966..16014,16692..18537))
 /gene="P0529E05.5"
 /complement(join(14141..14355,14614..14942,15117..15432,
 15672..15826,15966..16014,16692..18537))
 /gene="P0529E05.5"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB84390.1"
 /db_xref="GI:18461193"
 /translation="WTTIQLDPPVPAVARSATSCDHPDETFTGCTACURERLACL
 EASAAAPGKRSTSAIRSLFAPFVGAGGASVAGAGALPLRLRKSRCSCRGD
 ALAAAAAAGTARADRPORSCDVRSHSTLWLFPHDDGRVPSSSAADIVAPHQ
 QPPPPRPPIPDDELDDIIVYMHEDHIVYVPIVYDPSGIEETEPNVVARGA
 MKDHDFSSQPKKPTKDLKESFWVASVFSKMQKRRQKKEAAVAKAAAM
 PPEKPSKPSFLRRRLRKGASSELGASRCVDPFRSLDAGNSVDASWDEPR
 ASMDVYLFAGAGGIGLGRAPPLRLPVMVSLDTPATIVERDQIIPVEDADIDP
 PGSLQTRDYLDSTRRRLRLERSSVRPSFEYTERKPAPTTIANKESPLGSEF
 YHFIHAEDLDGRFSNGLIEDIASLEALSGSAAKPRRWKAMLMGFIRRAS
 GRSGGSPSDIADRSFSEAMPDLRAGVYVPMKQKQNSNLSARSPSNGGSRASH
 VDVNGSSARREHYLRNRSARYSPGRVNDGMLRYLTMRSGCGGGRARVAAA
 SETYKGTRLALYSMRCLALPLERYDLWSSADADTASHTMRPBRVYLDDEL
 ASIAVAVLAAHGNRPTQERTHSGIDRNDGSKDHTAGLSAPASAAFTKHDEP
 AGVAVPSVRQRRGRRAELGSTRYQWTRPACRPGPARLARTGRNRYSHVGR
 TSSRPCLGHAMVHQSGETHIRHGLINAGCCFSLPGVAMKARTGNQDITKTK
 QTSRTPPFSKMGALLIDOMARANCHRTIPTCAEOMQVHMSGMCCTVLSHG
 STVRKAMVAIVLAEKVCVERFDDISVVRPCH324021.24116..24188,
 24273..24483,24478..24651,24777..25855,25951..27372))
 /gene="P0529E05.6"
 /complement(join(23661..23843,23923..24021,24116..24188,
 24273..24483,24478..24651,24777..25855,25951..27372))
 /gene="P0529E05.6"
 /note="contains EST AU090537 (C1241)"
 /unknown protein"
 /codon_start=1
 /protein_id="BAB84391.1"
 /db_xref="GI:18461194"
 /translation="MAQRDEQILYDTAEGSSGYWNEEGCNPNODLNEGVARD
 AEGNQGEMERDVNGEESGSPQVQGRAGOGAAGKXGRIITEVEDGSR
 SAPDAKNTYRSHGVVDDVPSVNYWTRTRGHSFVPSSEKEMTLITLLETF
 TLPAETEKYRGTWKLMKAOEFOSEKELLYKTLKOTPTFDPFPLRLDMDFFVY
 KTGBOGAMMERKNENAKKRYHHHLSGGVSYVMPMESEATPDRGEPAANP
 ERSKPYVAGNCTLNPAADGIVEDQVPRBAARLTNAVBSAGSTPDRGEPAANP
 IQPHEHGOGRGKAVIPWKIGFKEDITHTYSQKSKDTEKADLEFRVSSYELNQ
 EVARVDERMARSNPPOTIPPAVSPGNSGASTQVGSQMDMWTQDEST
 CPVDDITQWTCLEHLPKSLIKVAGMAIPTDPGQTHCRPIPVGSKVEVELVVG

gene

CDS


```

repeat_region      4666..4689
/rpt_family="CTG)n"
repeat_region      5314..5607
/rpt_family="AluSg"
repeat_region      5690..5713
/rpt_family="TTTTG)n"
repeat_region      complement(5720..5995)
/rpt_family="AluSg"
repeat_region      complement(6032..6248)
/rpt_family="MIR"
repeat_region      complement(6274..6398)
/rpt_family="MER69A"
repeat_region      complement(6421..6597)
/rpt_family="MIR"
repeat_region      7452..7841
/rpt_family="LTRS2"
repeat_region      8208..8311
/rpt_family="MIR3"
repeat_region      complement(10123..10394)
/rpt_family="AluSx"
repeat_region      11628a..12039
/rpt_family="LTR32"
repeat_region      12041..12199
/rpt_family="L2"
repeat_region      12284..12362
/rpt_family="MIR"
repeat_region      14264..14312
/rpt_family="C-rich"
repeat_region      complement(14318..14603)
/rpt_family="AluSx"
repeat_region      complement(14955..15078)
/rpt_family="MIR"
repeat_region      15488..15638
/rpt_family="MIR"
repeat_region      complement(16167..16346)
/rpt_family="MIR"
repeat_region      16839..16983
/rpt_family="MIR3"
repeat_region      17119..17366
/rpt_family="MIR"
repeat_region      complement(17371..17533)
/rpt_family="L2"
repeat_region      17651..17756
/rpt_family="LIM4"
repeat_region      complement(17757..18054)
/rpt_family="AluSx"
repeat_region      18055..18494
/rpt_family="LIM4"
repeat_region      18495..18805
/rpt_family="AluSx"
repeat_region      18806..19007
/rpt_family="LIM4"
repeat_region      19009..19304
/rpt_family="LIM4"
repeat_region      19309..19304
/rpt_family="AluSx"
repeat_region      19289..19296
/rpt_family="AluSx"
/notes="Unresolved di-nucleotide repeat; restriction enzyme
digest fingerprint indicates
100 bp of missing sequence here"
complement(18297..19350)
/notes="<30 qual PCR product sequence only; single clone
coverage"
19305..20045
/rpt_family="LIM4"
complement(19531..20026)
/notes="PCR product sequence only"
complement(19619..19825)
/notes="<30 qual SNGF region"
repeat_region      complement(20049..20175)
/rpt_family="FLAM_C"
repeat_region      20176..20243
/rpt_family="LIM8"
repeat_region      20267..20315
/rpt_family="MER103"

```

```

repeat_region      20316..20452
/rpt_family="FLAM_A"
repeat_region      20453..20561
/rpt_family="MER103"
repeat_region      20768..20909
/rpt_family="LIM4"
repeat_region      complement(21428..21598)
/rpt_family="MIR"
repeat_region      23137..23194
/rpt_family="MIR"
repeat_region      complement(24411..24468)
/rpt_family="L2"
repeat_region      25356..25652
/rpt_family="AluSx"
repeat_region      25962..26003
/rpt_family="CT-rich"
repeat_region      26082..26135

```

Query Match 86.4%; Score 19; DB 9; Length 146101;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TCCACACACGACGACGACCT 21
Db 125454 TCCACACACGACGACGACCT 125436

RESULT 23
AP003250
LOCUS AP003250 146670 bp DNA linear PLN 29-MAR-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
PAC clone: P0443D08.
ACCESSION AP003250 BA000010
VERSION AP003250.3 GI:15408719
KEYWORDS

SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS
1 Saeki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
Katsuyose, Y., Wu, J., Nishimura, Y., Cheng, Z., Nagamura, Y.,
Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,
Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
Hamada, M., Harada, C., Hishitsuta, S., Honda, M., Ichikawa, Y.,
Idonuma, A., Iijima, M., Ikeda, M., Ikono, M., Itoh, S., Itoh, T.,
Itoh, Y., Itoh, Y., Iwabuchi, A., Kamiya, K., Karsawa, W., Katagiri, S.,
Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,
Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,
Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,
Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,
Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Teiji, K., Waki, K.,
Yamagata, H., Yamane, H., Yoshiki, S., Yoshitara, R., Yukawa, K.,
Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,
Yano, M., Jiang, J. and Gojobori, T.

TITLE The genome sequence and structure of rice chromosome 1
JOURNAL Nature 420 (6913), 312-316 (2002)
MEDLINE 22337376
PUBMED 12447438
2 (bases 1 to 146670)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: teasaki@nias.affrc.go.jp; URL: http://ryp.dna.affrc.go.jp/
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Aug 31, 2001 this sequence version replaced gi:14522861.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as Slicepredictor
(October 1998 version). The genomic sequence was searched against
NCBI Nonredundant Protein database, nr

COMMENT

(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP 2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DDBJ accession no. and RGP clone ID. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein. The orientation of the sequence is from 5' to 3' of the PAC clone. This sequence of P0443D08 clone has an overlap with OSJNA0083M16 (DDBJ: AP003214) at the position 1 to 624 and with P0480E02 (DDBJ: AP002913) clone at the position 141,467 to 146,670 of 3' end. The sequence of this clone starts at the position 138,088 of OSJNA0083M16 and ends at the position 5205 of P0480E02. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://xcp.dna.affrc.go.jp/genomeseg.html>.

FEATURES

source 1.146670
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="P0443D08"
2727..3327
/gene="P0443D08.1"
2727..3327
/note="contains ESTs D48568(S14863), AU089765(S14863) similar to 16.9 kDa heat shock protein
unknown protein"
/codon_start=1
/protein_id="BAB64123.1"
/db_xref="GI:15408720"
/translation="MSLYKLPPTLAFPDANPISFGTTVAADAMLASDTKSPANTYIE
SERIRAYVFRADLPAGYKKEVVEEGNVLVTGERSVREKKGSRSHIRSCA
TFGRFHLPPDDAVDLVPRASMDGMGLTVTPKVVTDKOPALAAAPVAVVAPVAVK
AIEASP"
complement(join(4071..4126,4283..4398,5525..5601))
/gene="P0443D08.2"
complement(join(4071..4126,4283..4398,5525..5601))
/gene="P0443D08.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB64124.1"
/db_xref="GI:15408721"
/translation="MGAMFIAKINPKFDNAKEFRVCAPOSTAGKACKLVKHHGEY
AGIQVNPSPSLMLTFPSREGVGLSATVRYLTH"
complement(6051..6500)
/gene="P0443D08.3"
complement(6051..6500)
/note="contains EST AU093701(C63333)"
/codon_start=1
/product="putative 16.9 kDa heat shock protein"
/protein_id="BAB64125.1"
/db_xref="GI:15408722"
/translation="MSLVRSNVDPDFPDNDPDDGVRSLVPATSRDRTAAFNANAV
DKETPESHVFRADLPYKKEVKEVEEGNVLVTSGRSKEKEDKDKMHRVERSSG
QPMRRRLPENAKVDQVAKMENGVLTVTPKAEVKEVKAIEISG"
complement(8660..10035)
/gene="P0443D08.4"
complement(8660..10035)
/note="Pseudogene
similar to non-LTR retroelement reverse transcriptase"
/pseudo
complement(10583..11035)

CDS
/gene="P0443D08.5"
complement(10583..11035)
/gene="P0443D08.5"
/note="contains EST AU092880(C60687)"
/codon_start=1
/product="putative 16.9 kDa heat shock protein"
/protein_id="BAB64126.1"
/db_xref="GI:15408723"
/translation="MSLVRSNVDPDFSLDMPDPSVRSVPATSDNDTAAFNANAR
IDMKETPESHVFRADLPYKKEVKEVEEGNVLVTSGRSKEKEDKDKMHRVERSS
QPMRRRLPENAKVDQVAKMENGVLTVTPKAEVKEVKAIEISG"
13625..14077
/gene="P0443D08.6"
13625..14077
/gene="P0443D08.6"
/note="contains EST AU092880(C60687)"
/codon_start=1
/product="putative 16.9 kDa heat shock protein"
/protein_id="BAB64127.1"
/db_xref="GI:15408724"
/translation="MSLVRSNVDPDFSLDMPDPSVRSVPATSDNDTAAFNANAR
IDMKETPESHVFRADLPYKKEVKEVEEGNVLVTSGRSKEKEDKDKMHRVERSS
QPMRRRLPENAKVDQVAKMENGVLTVTPKAEVKEVKAIEISG"
complement(join(16992..17439,18474..18547))
/gene="P0443D08.7"
complement(join(16992..17439,18474..18547))
/gene="P0443D08.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB64128.1"
/db_xref="GI:15408725"
/translation="MEGSHGQGVNTASHVVDVEREKPRPKYSFVTVIRLLCCRRK
KQADPVAAGGEDDGERSSILLTSRSSLELLVSDADDGAIIDAAVTCRNASLCAC
KSDQAPVAVLPFPPGLHPVWAPRAPHQWVTSFGSGRDAAGVCCRRRPFMRRLRLAM
RRWRPVLVAIEP"
complement(join(19682..20619,20791..21045,21998..22714,
23306..23926,24506..25205))
/gene="P0443D08.8"
complement(join(19682..20619,20791..21045,21998..22714,
23306..23926,24506..25205))
/gene="P0443D08.8"
/note="contains ESTs AU030772(E60216), AU030771(E60216) similar to wall-associated kinase
unknown protein"
/codon_start=1
/protein_id="BAB64129.1"
/db_xref="GI:15408726"
/translation="MSPSFFVYVSAWSIALMLAAARGAEESGGGCLIGSKCGDLN
ISSPFWIIQGOADKPCGPDIOYVCNNSGVATLSSSTSGFDIINIISGDDRTMVF
VHKLARLNTSGSISPVNTFALPIPTFISNSNHLVYNGCPAPAPCOOQGLIVE
TRCGNTPARIGRFRHGEDYDKYVEGSRSTYVLPLEBPDGAKNARVYELVAG
GELITWDLPPTSSAANRABOEGASATGCKTISPPVAVPEATRESCKW
LGQVICHADTPTLGYKRRYIQLIDITYGNNLSIVDHLGCPPIYSGSKSYSC
HFRNTTSKVGPSISTTNLNLPLFYSCKRLVPRDGGDVLVTRCKKTFARVGN
YSDSGDYPAFMEGCNATVPVGLDARSYEQILDGFLITWQETPSSASFLPAPA
SSSPGCLPTPCGKLTISYFMLEBGRPPCSFPOLKCNATGAYLTITTYEARVY
DIFGNTVAVVDENLPLATGCPAPFNSDGIWQAPFVISAANBELRSCKSLPA
AAAPGPHSLPCDDONSVRLVSDHLLHEDGIPGCPNFVYVIVORHNSMAGYIASM
RSGFLLEMAVSGDCKCGVSGGNCYSDDLEFACNCPGMHPCREBRKSEHACG
SSGILLIYISITFIAMHKKRKKOTRDKDMLHSSSMOSYSKIDELIGSSPHIYYEE
LEBNTDGFASRELDDGGTYTKKTLRGRVYAKRYKKNRYVEQGLNVDLISL
LHQNVLIVGCTSSSRDLVYETIPNGTVADHLDHGRAGEGLTFVRTTIAET
AALAVLHAEVLIHRDVKTNLILLNNFRVAVDGLSLFPLVTHVSTVQGTGPG
VAPVYHQCYKLTDSKDYVSFGVVLIELSSKAVMSRSHDINLANNALNRIQNHV
DOLVPEIYEETDSSTKRMVDVAAELAQCLQMDSESPPIREVEVLANCIKRGCPA
EKNKKAASPKEDSHLLKDSLOYSPDSVIRHFGSSTNSVANSNG"
complement(join(27648..27664,28123..28903))
/gene="P0443D08.9"
complement(join(27648..27664,28123..28903))
/gene="P0443D08.9"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB64130.1"

gene

CDS

/db.xref="GI:15408727"
/translation="MPTLLCLPLIAGLLI,CHRAACEPCATCNLTVPYPMILGAP
NFNSQSPSSALASCGHFAEYWGNGVSLASQILVLSIDYNSFVAHKEVA
DGGGVCRIDENISSLSLSPFTISSNRALCFYSCNGPEPIDGLVATISGCSK
PIYVAGIYDRNDPAIKAGCTCYSLPLMPLMSPANLGNAGNYSPCFFKGVLEMO
KNRGDDCDACNGSGGCRITINDSAAACACUSDCRKRSTCRGCRST"
complement(15408727) .29938,29938, .30738,30790, .30937,
30978, .31158,31206, .31362,31389, .31517,31689, .31800,
31910, .32176)
/gene="P0443D08.10"

Query Match 86.4%; Score 19; DB 8; Length 146670;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 24
AP003442 151703 bp DNA linear HTG 29-MAR-2003
LOCUS Oryza sativa (japonica cultivar-group) chromosome 1 clone B1096A10,
DEFINITION *** SEQUENCING IN PROGRESS ***
ACCESSION AP003442.1 GI:13486911
VERSION AP003442.1
KEYWORDS HTG, HTGS, PHASE2.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehharctoidae; Oryzaceae; Oryza.

REFERENCE
AUTHORS Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
Katayose, Y., Wu, J., Nishimura, Y., Cheng, Z., Nagamura, Y.,
Antonio, B.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,
Chuden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
Hamada, M., Harada, C., Hishita, S., Honda, M., Ichikawa, Y.,
Iiduma, A., Iijima, M., Ikeda, M., Ikono, M., Itoh, S., Itoh, T.,
Itoh, Y., Iwabuuchi, A., Kamiya, K., Karasawa, M., Katagiri, S.,
Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,
Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,
Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,
Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shirokawa, T.,
Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,
Yamagata, H., Yamane, H., Yoshiki, S., Yoshimura, R., Yukawa, K.,
Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J.H., Kim, H.T., Eun, M.Y.,
Yano, M., Jiang, J. and Gojobori, T.
The genome sequence and structure of rice chromosome 1
Nature 420 (6913), 312-316 (2002)

TITLE
JOURNAL
MEDLINE
22337376
12447438
2 (bases 1 to 151703)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasa@kenias.affrc.go.jp, URL:http://rsp.dna.affrc.go.jp/
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT

* NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 151703: contig of 151703 bp in length.
Location/Qualifiers
1..151703
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="B1096A10"

BASE COUNT 33379 a 36328 c 37093 g 38853 t 50 others
ORIGIN

Query Match 86.4%; Score 19; DB 2; Length 151703;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GCCACCCACGACGACCA 22
DB 16614 GCCACCCACGACGACCA 16632

RESULT 25
AL450333/c 156964 bp DNA linear PRI 06-APR-2001
LOCUS Human DNA sequence from clone RP11-2506 on chromosome 6, complete
DEFINITION sequence.
ACCESSION AL450333 AC055814
VERSION AL450333.13 GI:13561015
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
Katayose, Y., Wu, J., Nishimura, Y., Cheng, Z., Nagamura, Y.,
Antonio, B.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,
Chuden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
Hamada, M., Harada, C., Hishita, S., Honda, M., Ichikawa, Y.,
Iiduma, A., Iijima, M., Ikeda, M., Ikono, M., Itoh, S., Itoh, T.,
Itoh, Y., Iwabuuchi, A., Kamiya, K., Karasawa, M., Katagiri, S.,
Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,
Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,
Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,
Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shirokawa, T.,
Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,
Yamagata, H., Yamane, H., Yoshiki, S., Yoshimura, R., Yukawa, K.,
Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J.H., Kim, H.T., Eun, M.Y.,
Yano, M., Jiang, J. and Gojobori, T.
The genome sequence and structure of rice chromosome 1
Nature 420 (6913), 312-316 (2002)

COMMENT

* NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.


```

repeat_region 65383..65748
                /note="THE1B repeat: matches 1..364 of consensus"
repeat_region 65757..67368
                /note="THE1B-INTERNAL repeat: matches 3..1637 of consensus"
repeat_region 67371..67738
                /note="THE1B repeat: matches 1..364 of consensus"

Query Match      86.4% Score 19; DB 9; Length 156964;
Best Local Similarity 100.0%; Pred. NO 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy              4 GGCACCACCGACGACCA 22
Db              31632 GGCACCACCGACGACCA 31614

RESULT 26
AC012132/c
LOCUS
DEFINITION Homo sapiens clone RP11-5116, WORKING DRAFT SEQUENCE, 4 unordered
            pieces.
ACCESSION AC012132
VERSION AC012132.5 GI:11079551
KEYWORDS HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-5116
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163990)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckery,F., Boguslavsky,L., Boukhalter,B.,
            Brown,A., Caselle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., Dextralano,K., Dewar,K., Domino,S., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Hewland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lechoczky,J., Lien,C., Locke,K., MacDonald,P., Marcus,N.,
            McMan,P., McGuck,A., McKernan,K., McLaughlin,J., Meldrum,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Tommann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Teste,S., Tirkell,A., Vassiliou,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Nov 3, 2000 this sequence version replaced g11966279.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L2859
            Center clone name: 5.1.16
            ----- Summary Statistics
            Sequencing vector: M13, M77815, 49% of reads
            Sequencing vector: Plasmid, n/a, 51% of reads
            Chemistry: Dye-terminator Big Dye, 100% of reads
            Assembly program: Phrap, version 0.960731
            Consensus quality: 162675 bases at least Q40
            Consensus quality: 163217 bases at least Q30
            Consensus quality: 163480 bases at least Q20
            Insert size: 163000; agarose-efp
            Insert size: 163690; sum-of-contrigs

```

```

Quality coverage: 8.8 in Q20 bases; agacrose-fp
Quality coverage: 8.8 in Q20 bas.
NOTE: This is a 'working draft' sequence. It currently
consists of 4 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
*
* 1 4577: contig of 4577 bp in length
*
* 4578 4677: gap of 100 bp
*
* 4678 5865: contig of 1188 bp in length
*
* 5865 5965: gap of 100 bp
*
* 5965 24687: contig of 18722 bp in length
*
* 24688 24787: gap of 100 bp
*
* 24788 163990: contig of 139203 bp in length.
*
Location/Qualifiers
1. 163990
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:3606"
/clone="RP11-5116"
/clone_11b="RPC1-11 Human Male BAC"
1. 4577
/misc_feature
/note="assembly_fragment
clone end:SP6
vector side:left"
4578_5865
/misc_feature
/note="assembly_fragment"
5965_124687
/misc_feature
/note="assembly_fragment"
24788_163990
/misc_feature
/note="assembly_fragment
clone end:T7
vector side:right"
BASE COUNT 54038 a 30846 c 30609 g 48136 t 301 others
ORIGIN
Query Match 86.4%; Score 19; DB 2; Length 163990;
Best Local Similarity 100.0%; Freq. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 4 GCCACGACGACGACGACCA 22
Db 145332 GCCACGACGACGACGACCA 145314
RESULT 27
AC027801 164914 bp DNA linear PRI 08-OCT-2001
LOCUS Homo sapiens chromosome 17, clone R11-304F15, complete sequence.
AC027801
AC027801.10 GI:15983571
FEATURES
FTS.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 164914)
1. Birren,B., Linton,L., Nussbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone R11-304F15
Unpublished
2 (bases 1 to 164914)
1. Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Chepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArillano,K., Dewar,K., Diaz,T.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Illey,I., Johnson,R., Jones,C., Kann,L., Karakas,A.,

```

```

Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieb,C., Liu,G., Locke,K., MacDonald,P., Margolis,N.,
McCarthy,M., McEwen,P., McGuirk,A., McKernan,K., McPheters,R.,
Melgrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenda,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.K., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 164914)
Birren,B., Linton,L., Nusbaum,C., Landers,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bassien,V., Boguslavsky,L., Bouckgalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gaidyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heatford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Margolis,N., Matthews,C.,
McCarthy,M., McEwen,P., McKernan,K., McPheters,R., Melgrim,J.,
Meneus,L., Mihova,T., Mlenda,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strause,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topfham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-QCT-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 8, 2001 this sequence version replaced g1:3493143.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9064
Center clone name: 304_F_15
-----
Location/Qualifiers
1..164914
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-304F15"
/clone_lib="RP11-304F15 Human Male BAC"
/complement(662..767)
repeat_region
/rpt_family="L2"
/complement(1078..1623)
repeat_region
/rpt_family="L2"
/complement(2818..3069)
repeat_region
/rpt_family="AluSx"
/complement(3238..3514)
repeat_region
/rpt_family="AluSg"
3625..3836
/rpt_family="L1P2"
repeat_region

```

```

repeat_region 4094..4491
/rpt_family="AluDb"
repeat_region 4493..4800
/rpt_family="AluY"
repeat_region 4808..5110
/rpt_family="AluSx"
repeat_region 5274..5362
/rpt_family="L2"
repeat_region 6664..6915
/rpt_family="L2"
repeat_region 6960..7289
/rpt_family="L2"
/complement(7467..7631)
repeat_region
/rpt_family="MIR"
7632..7937
/rpt_family="AluJc"
/complement(7938..8013)
repeat_region
/rpt_family="MIR"
8884..9198
/rpt_family="AluSp"
9208..9512
/rpt_family="AluJb"
/complement(12069..12184)
/rpt_family="L2"
/complement(12606..12809)
/rpt_family="AluSg"
12866..13158
/rpt_family="MER46C"
13789..14061
/rpt_family="AluSg"
/complement(13957..13969)
/note="<30 qual SNGL region"
14705..14731
/rpt_family="TTG)n"
14816..15074
/rpt_family="T1ggers5"
/complement(14954..14972)
/note="<30 qual SNGL region"
/complement(15000..15005)
/note="<30 qual SNGL region"
15059..15224
/rpt_family="T1ggers5"
/complement(15225..15525)
/rpt_family="AluJc"
15526..15610
/rpt_family="T1ggers5"
/complement(15987..16283)
/rpt_family="AluSx"
/complement(16482..16701)
/rpt_family="L2"
/complement(16702..16998)
/rpt_family="AluSx"
/complement(16999..17092)
/rpt_family="L2"
17421..17442
/rpt_family="AT_rich"
17511..17542
/rpt_family="AT_rich"
17547..17705
/rpt_family="L1MA6"
/complement(18734..18977)
/rpt_family="MIR"
19009..19040
/rpt_family="AT_rich"
/complement(19465..19775)
/rpt_family="AluY"
19970..20264
/rpt_family="MER102a"
/complement(21790..22074)
/rpt_family="AluSg"
/complement(22075..22380)
/rpt_family="AluJc"
/complement(22424..22723)
repeat_region

```


JOURNAL Submitted (23-NOV-2002) Stanford Human Genome Center, 975 California Avenue, Palo Alto, CA 94304, USA

REFERENCE 4 (bases 1 to 172650)

AUTHORS Grimwood,J., Dickson,M., Schmutz,J., Stuart,A., Miyake,T., Amemiya,C. and Myers,R.M.

TITLE Direct Submission

JOURNAL Submitted (04-DEC-2002) Stanford Human Genome Center, 975 California Avenue, Palo Alto, CA 94304, USA

COMMENT On Dec 4, 2002 this sequence version replaced gi:24942881.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the Stanford Human Genome Center (<http://www-shgc.stanford.edu>) and the Virginia Mason Research Center (http://vmresearch.org/lab_research/default.htm)

Quality: Phrap Quality >=40 100% Of Sequence;

Estimated total Number of Errors is 0.1.

FEATURES

source

1. 172650

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="RP43-93C21"

BASE COUNT 41253 a 45062 c 44566 g 41779 t

ORIGIN

Query Match 86.4%; Score 19; DB 9; Length 172650;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3 TGCCACCCAGCAGCACC 21

56450 TGCCACCCAGCAGCACC 56468

DB

AC126925 176258 bp DNA linear HTG 06-JUN-2003

AC126925/c

LOCUS

DEFINITION

Canis familiaris clone RP81-332E11, WORKING DRAFT SEQUENCE, 12

ORDERED pieces.

AC126925

AC126925.6 GI:31442444

KEYWORDS

HTG; HTGS_PHASE2; HTGS_DRAFT.

SOURCE

Canis familiaris (dog)

ORGANISM

Canis familiaris

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE 1 (bases 1 to 176258)

AUTHORS Antweiler,A., Aylele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Carlsaga,K., Chu,G., Coleman,B., Coleman,H., Engle,J., Granter,S., Guan,X., Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Lalic,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masciello,C., McKerril,B., McDowell,J., Pasquigian,C., Pearson,R., Portroy,M.E., Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K., Sisson,C., Stantipod,S., Thomas,J.W., Thomas,P.J., Tespouri,V., Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.

TITLE NISC Comparative Sequencing Initiative

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 176258)

AUTHORS Green,E.D.

TITLE Direct Submission

JOURNAL Submitted (10-JUN-2002) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA

3 (bases 1 to 176258)

REFERENCE Green,E.D.

TITLE Direct Submission

JOURNAL Submitted (06-JUN-2003) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA

On Jun 6, 2003 this sequence version replaced gi:28209436.

COMMENT

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_zoengrl@nih.gov

----- Project Information

Center project name: cwp

Center clone name: 332E11

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 173760 bases at least Q40

Consensus quality: 174423 bases at least Q30

Consensus quality: 174916 bases at least Q20

Insert size: 152000; agarose- $\frac{1}{2}$

Insert size: 175158; sum-of-contigs

Quality coverage: 17,46x in Q20 bases; agarose- $\frac{1}{2}$

Quality coverage: 15.15x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and the accession number will be preserved.

1

11425: contig of 11425 bp in length

11426 11525: gap of unknown length

11526 27554: contig of 16029 bp in length

27555 27654: gap of unknown length

27655 29974: contig of 2320 bp in length

29975 30074: gap of unknown length

30075 36241: contig of 6167 bp in length

36242 36341: gap of unknown length

36342 77002: contig of 40661 bp in length

77003 77102: gap of unknown length

77103 117018: contig of 39916 bp in length

117019 117118: gap of unknown length

117119 119041: contig of 1923 bp in length

119042 119141: gap of unknown length

119142 168386: contig of 39247 bp in length

168387 158389: gap of unknown length

158389 158489: contig of 9545 bp in length

158489 168133: gap of unknown length

168134 170716: contig of 2583 bp in length

170717 170816: gap of unknown length

170817 174429: contig of 3613 bp in length

174430 174528: gap of unknown length

174529 176258: contig of 1729 bp in length.

176259

Location/Qualifiers

1. 176258

/organism="Canis familiaris"

/mol_type="genomic DNA"

/db_xref="taxon:9615"

/clone="RP81-332E11"

/clone_1b="RP81"

1. 86033

/note="clone overlaps with GenBank Accession Number AC126237 clone RP81-414022 (center project name cwp)"

1. 11425

misc_feature

misc_feature

misc_feature

```

/misc_feature /note="assembly_fragment
clone_end:SP6
vector_side:left"
11526..27554
/note="assembly_fragment"
27655..29974
/note="assembly_fragment"
30075..36241
/note="assembly_fragment"
36342..77002
/note="assembly_fragment"
77103..117018
/note="assembly_fragment"
117119..119041
/note="assembly_fragment"
119142..158388
/note="assembly_fragment"
158489..168033
/note="assembly_fragment"
168134..170716
/note="assembly_fragment"
170817..174429
/note="assembly_fragment"
174530..176258
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 41730 a 44533 c 46492 g 42401 t 1102 others
ORIGIN

```

```

Query Match 86.4%; Score 19; DB 2; Length 176258;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 172121 GCCACCACGACGACCA 172103

```

```

RESULT 31
LOCUS 177793 bp DNA linear INV 05-MAY-2001
DEFINITION Drosophila melanogaster, chromosome X, region 15B-15F, BAC clone
AC012096
AC012096 complete sequence.
AC012096.8 GI:13957595
HTG.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 177793)
Cejnifer,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amentides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Gary,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Idegam,C., Jatali,M., Kruse,D., Li,P., Mattei,B., Mostrel,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Palacio,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Seaplerton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome X, region 15E-15F
Unpublished
2 (bases 1 to 177793)
Cejnifer,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazek,R.G.,
Buenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

```

```

Kim,E., Lee,B., Lewis,S., Li,P., Lomutan,M.A., Mazda,P.,
Mostrefi,A.R., Mostrefi,M., Nixon,K., Pacלב,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Seth,H., Shit,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (20-Oct-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On May 5, 2001 this sequence version replaced gi:6957581.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdgs@fruitfly.berkeley.edu.
Location/Qualifiers
1..177793

```

FEATURES

```

source
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y1 cn bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/map="15E-15F"
/clone="BACR18002 (D1111)"
/clone_id="RPCL-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6)"
BASE COUNT 51373 a 37606 c 37116 g 51698 t
ORIGIN

```

```

Query Match 86.4%; Score 19; DB 3; Length 177793;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 78693 GCCACCACGACGACCA 78711

```

```

RESULT 32
LOCUS 181062 bp DNA linear HTG 22-MAY-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-655J14 map 2, WORKING DRAFT
AC062006
AC062006 26 unordered pieces.
AC062006.2 GI:8018045
HTG; HTGS PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 181062)
Birken,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 2, clone RP11-655J14
Unpublished
2 (bases 1 to 181062)
Birken,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,E.,
Boguslavsky,L., Bouhassal,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choquet,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dekrellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Gallagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Headford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kam,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levin,J., Lieou,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meldrim,J., Meneus,L., Minova,T., Miranda,C., Mlenga,V., Morrow,J.,

```



```
misc_feature /note="assembly_fragment"
misc_feature 116401..124607
             /note="assembly_fragment"
misc_feature 124708..135397
             /note="assembly_fragment"
misc_feature 135498..147379
             /note="assembly_fragment"
misc_feature 147680..162430
             /note="assembly_fragment"
misc_feature 162531..181062
             /note="assembly_fragment"
BASE COUNT 48952 a 39861 c 40075 g 49650 t 2514 others
ORIGIN

Query Match      86.4%  Score 19;  DB 2;  Length 181062;
Best Local Similarity 100.0%;  Pred. No. 21;
Matches 19;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY      4  GCCACCCAGCAGACACCA 22
Db      25072  GCCACCCAGCAGACCA 25090

RESULT 33
AL646084/c
LOCUS      AL646084      215050 bp      DNA      linear      BCT 02-SEP-2002
DEFINITION Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
ACCESSION  AL646084 AL646053
VERSION     AL646084.1 GI:17431752
KEYWORDS
SOURCE
ORGANISM    Ralstonia solanacearum
             Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
             Burkholderiaceae; Ralstonia.
REFERENCE
AUTHORS     1 Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
             Ariat,M., Billault,A., Broctier,P., Camus,J.C., Catcolico,L.,
             Chanderli,M., Cloisne,N., Claudel-Renard,C., Cumnac,S., Demange,N.,
             Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,E.,
             Signier,P., Thebaud,P., Whalen,M., Winkler,P., Levy,M.,
             Weissenbach,J. and Boucher,C.A.
             Genome sequence of the plant pathogen Ralstonia solanacearum
             Nature 415 (6871), 497-502 (2002)
JOURNAL     MEDLINE
PUBMED      21681879
REFERENCE   2 (bases 1 to 215050)
AUTHORS     Boucher,C.A.
TITLE       Direct Submission
JOURNAL
SUBMITTED   (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
             Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
             Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
             BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
             Dausset-CBPH, 27 rue Juliette Dodu, 75010 Paris, France, LMWG CNRS
             118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
             URV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
             Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
             F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
             INRA, BP27, F31326 Castanet-Tolosan Cedex
             Christian.Boucher@toulouse.inra.fr
             http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
SOURCE
1..215050
/organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
/complement(61..984)
/gene="RS05325"
/note="synonym: RSP1280"
/complement(61..984)
/gene="RS05325"

gene
CDS
/EC_number="2.7.1.-"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence: probable"
Gene name confidence: hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE SUCAR KINASE PROTEIN"
/protein_id="CAD18431.1"
/db_xref="GI:17431753"
/db_xref="SPTREMBL:Q8XOE7"
/translation="MLTNNLPQLVFGESALTDIFREADGKRWVSAGSCNNVANVCA
RLGAPTFGAGTVSRDVGDELTALSRQGLDKRFMRQVAPRLPAMVPAHPPOYEFI
GENSADLAFDNGVLPDGMIBAAQIVHPSLSIARQPLALETAVAAKRIAFD
PYRDAMAGDPYRPTLRMAAGIASYIKISGSDIRGLPEIDETSLLAQDAAPTAI
LMTRAAGKTLITPEGTLFQPLPTIVADITGAGDSMGMILSLITPEAMPASHLE
PSAASAAVCAQAGAAVPTRETVSALAVSR"
/complement(1270..6378)
/gene="RS05326"
/note="synonym: RSP1281"
/complement(1270..6378)
/gene="RS05326"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence: hypothetical"
Gene name confidence: hypothetical
Predicted by Codon usage
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHEICAL PROTEIN"
/protein_id="CAD18432.1"
/db_xref="GI:17431754"
/db_xref="SPTREMBL:Q8XOE8"
/translation="MLTGLISRSKASPSDLAVRRKPLAEPGSPKAGTRESNSLQ
OLFDDPNVAVSAGAAVOYOPBAGMSGLADGASVAPSGARTVTAASAIHEI
TPVPSPAIVPEHAPGAPASAAAASGAVAGCVPLRREPPLDAFKKXKTVARHP
ADRYQAHOLELRLNRYVSLTWNATGGLDIDYKGRVYVGMHSGSLRAIVL
FAHPMKQKCSPPIGLNLANSISPDGLWEHGOQLQWDTQKQMPHALPGMRA
PITLGRQIDENAVRAGSMILKXGAGGIRAYEVAGLEFSAVNDAGRLADGEG
RLCRDPSAOPRPPIQLADGAPAFEFVQKGPKEPIQLARADFVALPBGHITLFAH
DOGHLYOGDHLAADPAAGEIKARVGVVPMGSGADMGVETLATSPEGDKALFH
AAGRSBGQRYTLAAMDGOQWOPMHWBQGLILMSRGLQAPAMRYVATCTYNGAALIS
AAGASQDQGNHWRPLIQADCTPLADRLDKLPGLADAKPVYALOSTPAGSGLI
ALDGGNNARLPARGVAAVAPQRYVSQSVLAAASAPTRDRAVDGAGVAYHITDD
HALMRTPPGQOPQLPAPPGQARLEQIVASSDQOVFLAPSPDGGAAQLSLRY
DHASASWADCOAGLSVLDPEVRLGISLGLQLITGGADEQRTHRVAPLQKRY
RLATGTEPTVSOALTNRPRVRIPTGVATLHRAGGVTERVOSNFTVAQHVH
RLAMPKAVNRAEHLNKRDMABEYDMKATLELEPLLHHRMWPLPPTVAAGPVA
APVPEATPAQCALGDAIDKMLGALQIVARVAGPDLTLAAAGRRORVAPAKA
KPPDPLILPKMSMLDEVRPAIPEAKQASPPAKAPVVDPRALIGRHLHIVELMT
LSEHVXKLPADLNGQRTRRRPLILGAVNRHLITDLARIVTLPTVAVISDSA
VEVARQSVETKRVLAAPFSGSNWMDAEAFENIOTFKKRWKQSPFNKRFASSHAV
RDASSPAMKAPFAMKGSNSRSTLPEIKSGASAGVTSAGVLPKRVAPFRLQGG
GVDRRTWAGVERITADKLAEGPLVAFPNOSKSGLTMGGMGMDERPKLVGFRLQGG
QAASVAVHGGKAMLVAPDNIDEFARLPDPASHDQRLLELNGQALIDMI EQG
LNLNANAGLIGVADKIPGFAPARHGGSGTSAFOQGLQRGVAGAVNMGVDFEL
KLQAHMERISGEYVGGGKANAAPASIVQGGALPLPHYSADFTVLRLNITLIGASY
ELSGVESPKRTLLDMKTAARVYTPOMSSQALAAAREVPSOALGHGDEPLKILITTLQ
AAKSTWARTTHERASVYDRAEQLLDQCLASGRNMLPGAKTFNIPNRSIYDTR
KNSKARSLGLKFAEAREPAPVGLADAMKASERDGVNVRVFCMDPSYINAVNR
LMBGRKLSVAFNMTWARTVPAVRYLTEICADSSNSRGAFTLNDPLPLAFDSAEVSR
SLFAEVLARVGLNGRLGADLPLGAQRAVAGQCVLSPFDVAGVPPAAGAAPAQA
PDLPLQSQSLPSERPAGPALKARSLSRSXLE"
/complement(7105..9657)
/gene="RS05327"
/note="synonym: RSP1282"
/complement(7105..9657)
/gene="RS05327"
/EC_number="2.7.3.9"
```



```
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MULTIFUNCTIONAL PROTEIN :
PHOSPHOCARRIER PROTEIN HPR (PROTEIN H) AND
PHOSPHENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE"
/protein_id="CAD18433.1"
/db_xref="GI:17431755"
/db_xref="SPRMBL:08XOE5"
/translation="MSSRQLQKSLAPLSGRVPIEDVDPAPFQKRWGGLSDIPAT
DVLAIPIDRVIDFDHRAHAIYADAGVEIWHVGLDVTVMAGGFEALVKGDTVR
ACQPLRFPALVAARASALTEIVVNGSRVQMDKASVAGSVALTLLEADY
ADVAGDAPGAGCEALLOSPPIVLPNPAHLHARPAARLASRGPAAITLTVGQQA
DAKSAVAMALATRCGEVRLSGSGSDAQAALADILRSGSGSAGTALPOKKA
APALHRSPPVEEPAPAGSLRGVAGPGVAVOIVRMSGIAPENGTGPEPEOEARL
HALQTAASQIKSLVYADTLPPTPQALINDAHLEPDMLEAARACGRASAGA
WILACQSVBQHLHAPQVREPADVROGVRLVITGAEFRSLPERSTIYAD
LTPSDITIMDRSKVAGLCTVSGGPTSHVALIARSGIIPAVCGVPAQALSIRGTAVL
DGTAGVLCADPDPALRAEVRNLAARQLODROTARHPRTGRHGEVEVANVRA
AEAREVAAGGEGVGLRSEBPLEFEDRTAPDEDOAAVAQAVALGEPPLIYTLQ
IGGDKPLRLVPLEKEDNPFLGRIGRVS.LAYELFRSOLRAMLRAPGNGIINPMV
SDDEVILVAKRLIREQARVPIPVKIGWILEVPAVAIYEAALGAVDPGSGINDILO
XTLAIRGHAGLAARVDAHNPVLRMIATVTGGAHAGKMGVCANLSDLAAYVLV
GLGVDELIVSVPIASVXALSRDEDCRALARKVLPUGSARVREALAQ"
complement(9714..10958)
/ gene="RS05328"
/ note="synonym: Rsp1283"
complement(9714..10958)
/ gene="RS05328"
/function="cell processes; transport of small molecules;
carbohydrates, organic acids, alcohols"
/ note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE PORIN TRANSMEMBRANE PROTEIN"
/protein_id="CAD18434.1"
/db_xref="GI:17431756"
/db_xref="SPRMBL:08XOE4"
/translation="MKKTIITLIIAPWALTEAMAVNLGEGLEFNQYLPMGVSEI
QNDAGSNKALASGAEFLRINEGDTIEASLKVFQPLPGITAAAFCTGVNGSLA
NRGRTDQGMVTKKAYITTSGRASPAITFWAGKRIEREVEVITLHFFSVGGPTI
DTGVGATNIPVEQARLGIVSRSGRSPPTNRDARLNDLVINIEVTRDRLRI
GEWMGEFPGNGGALITPKXSPNFPLEPSVNSVVOGSLTVALNIEVTRDRLRI
TNSVRLIDVNMVQVGRFGQALAVYORGLNHSOGSGTGSVGSVAIVPYKLAS
EVALTLRVNGSLGRIDKRYTVAALASGPGILSRPEIRIYTRVAMNDALAAQAG
WGMSAGRSSTNLVGIQLESWM"
complement(11009..12430)
/ gene="RS05329"
/ note="synonym: Rsp1284, scrB"
complement(11009..12430)
/ gene="RS05329"
/ GC number="3.2.1.26"
/function="small molecule metabolism; central intermediary
metabolism; misc. glucose metabolism"
/ note="Product confidence : putative
Gene name confidence : putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
```

```
Query Match 86.4%; Score 19; DB 1; Length 215050;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 4 GCCACACAGCAGCAGCAGCA 22
Db 125862 GCCACACAGCAGCAGCA 125844
```

```
RESULT 34
AC126296/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-53A13, WORKING DRAFT SEQUENCE, 2
uncloned pieces.
```

```
AC126296.3 GI:30580982
KEYWORDS
HTG; HTG_PHASE1; HTG_DRAFT; HTGS_FULFILLOR.
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
```

```
REFERENCE
AUTHORS
Murray D. Nattie, Metzker M. Lee, Abramson S., Adams C., Alder J.,
Allen C., Allen H., Albrooke S., Amin A., Arguano D.,
Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benhmed F.,
Blawie K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryan N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavaros I., Caesar H., Center A.,
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
Davila M. L., Davis C., Davy-Carroll L., De Anda C., Deckerich D.,
Delgado O., Denson S., Deramo C., Ding Y., Dink H., Diya K.,
Diaper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Davies K.,
Egan A., Escotto M., Eugene C., Evans C. A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Fraser C. M., Gabis A., Ganta R., Garcia A., Garner T., Garza M.,
Gebrgeorgis E., Geier K., Gill R., Grady L., Guerra M., Guevara W.,
Gunaratne P., Haaland M., Hamill C., Hamilton C., Hamilton K.,
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
Hernandez R., Hines S., Hladun S. L., Hodgson A., Hoques M.,
Hollins B., Howells S., Hulik S., Hume J., Idelbird D., Jackson A.,
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
Kasparov S., Kelly S., Kelly S., Khan Z., King L., Kover C.,
Kowis C., Kraft C. L., Leonard H., Levay J., Lewis L., Li Z., Liu J.,
Liu Y., Liu Y., Liu Y., London P., Longacre S., Lopez J.,
Loenenhuwa L., Louisedge H., Lozado R. J., Lu X., Ma J.,
Maheshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangum A.,
Mangum B., Mapa P., Martin K., Martin R., Martinez E.,
Mawhney S., McLeod M. P., McNeill T. Z., Mennen E.,
Mlosoevic A., Miner G., Minja E., Montemayor J., Moore S.,
Morgan M., Morris K., Morris S., Munidasa N., Murphy M., Nair L.,
Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,
Nwokeneme O., Okwuonu G., Olarunsgoon A., Pal S., Parks K.,
Pasternak S., Paul H., Perez A., Perez L., Plankoch C.,
Plopper F., Polindexter A., Popovic D., Primus E., Pu L., L.,
Puzo M., Quiroz J., Rachlin E., Reeves K., Regier M. A., Reigh R.,
Riley B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S. J.,
Sanders N., Saverly G., Scherer S., Scott G., Shatman S., Shen H.,
Shetty J., Shvartsbeyn A., Sisson I., Sitter C. D., Sma's D.,
Sneed A., Sodergren E., Song X. Z., Sorelle R., Sosa J.,
Steimle M., Strong R., Sutton A., Svatek A., Taber P., Taylor C.,
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Umanli K.,
Valas R., Vera V., Villanana D., Waldron L., Walker B., Wang J.,
```

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 234920)
Worley, K.C.

Direct Submission
Submitted (05-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 234920)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23269045. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GKX2
Center clone name: CH230-53A13

----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 22413 bases at least Q40
Consensus quality: 225915 bases at least Q30
Consensus quality: 227447 bases at least Q20
Estimated insert size: 233973; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 180394: contig of 180394 bp in length
* 180395 180494: gap of unknown length
* 180495 234920: contig of 54426 bp in length.
Location/Qualifiers
1. 234920
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-53A13"
180495..181590
/notes="wgs_contig"
184044..185385
/notes="wgs_contig"

BASE COUNT 68124 a 51943 c 59758 t 6214 others

ORIGIN
Query Match 86.4%; Score 19; DB 2; Length 234920;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3 TGCCACGACGACGACGAC 21
|||||
Db 47373 TGCCACGACGACGAC 47355

AC095275 245660 bp DNA linear HTG 09-NOV-2002
Rattus norvegicus clone CH230-11020. *** SEQUENCING IN PROGRESS
***, 6 unordered pieces.
AC095275
AC095275.8 GI:24818045
HTG, HTGS, PHASE1, HTGS DRAFT, HTGS ENRICHED.
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 245660)
Munry, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, N., Barnstead, M., Benahmed, F., Bhatnagar, K., Blair, J., Blankenship, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Dreier, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Healand, W., Hamill, C., Hamilton, C., Hamilton, C., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hu, Y., Hume, J., Idler, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenz, W., Loulsegged, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahoud, M., Mallory, K., Mangum, A., Mangum, B., Mapu, P., Martin, K., Martin, K., Martinez, E., Maloney, S., McLeod, M.P., McNeill, T.Z., Meenan, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Narkervic, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokoleh, O., Okunolu, G., Olarnunaggon, A., Pal, S., Parks, K., Paeternak, S., Paul, H., Perez, A., Perez, L., Frankoch, C., Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Kelly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sison, I., Sitter, C.D., Smaiz, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sos, J., Steinle, A., Strong, R., Sutton, A., Svatek, A., Tabore, F., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villabona, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,

```

TITLE      Weinstein, G. and Gibbs, R.A.
JOURNAL    Direct Submission
REFERENCE  Unpublished
AUTHORS    2 (bases 1 to 245660)
           Worley, R.C.
JOURNAL    Direct Submission
           Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
           3 (bases 1 to 245660)
           Rat Genome Sequencing Consortium.
           Direct Submission
           Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
           On Nov 9, 2002 this sequence version replaced gi:23264505.
           The sequence in this assembly is a combination of BAC based reads
           and whole genome shotgun sequencing reads assembled using Atlas
           (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
           in the feature table below represents a scaffold in the Atlas
           assembly (a 'contig-scaffold'). Within each contig-scaffold,
           individual sequence contigs are ordered and oriented, and separated
           by sized gaps filled with Ns to the estimated size. The sequence
           may extend beyond the ends of the clone and there may be sequence
           contigs within a contig-scaffold that consist entirely of whole
           genome shotgun sequence reads. Both end sequences and whole genome
           shotgun sequence only contigs will be indicated in the feature
           table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDTY
Center clone name: CH230-11J20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 222455 bases at least Q40
Consensus quality: 226178 bases at least Q30
Consensus quality: 228871 bases at least Q20
Estimated insert size: 233816; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 117281: contig of 117281 bp in length
* 117282 117381: gap of unknown length
* 117382 230056: contig of 112675 bp in length
* 230057 230156: gap of unknown length
* 230157 240261: contig of 10105 bp in length
* 240262 240361: gap of unknown length
* 240362 242161: contig of 1800 bp in length
* 242162 242261: gap of unknown length
* 242262 243988: contig of 1727 bp in length
* 243989 244088: gap of unknown length
* 244089 245660: contig of 1572 bp in length.
Location/Qualifiers
1..245660
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-11J20"
2131..2765
/note="clone_boundary"

misc_feature
113919 GCCACGACGACGACGACCA 22
Db 113919 GCCACGACGACGACGACCA 113937

Query Match 86.4%; Score 19; DB 2; Length 245660;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: August 15, 2003, 09:34:14
Job time : 612.525 secs

BASE COUNT 71706 a 43859 c 43188 g 71170 t 15737 others
ORIGIN
misc_feature
230157..231407
/note="wgs contig"
clone_end:Sp6
site:EcoRI
end_sequence:BH342597"

```